AD 617118

AD

Best Available Copy

TR-1295

A MULTIPLE TESTING OF THE ABC METHOD AND THI DEVELOPMENT OF A SECOND GENERATION MODEL

PART 1
PRELIMINARY DISCUSSIONS OF METHODOLOGY

by

Berthold Altmann

SUPPLEMENT
COMPUTER PROGRAMS OF THE HDL INFORMATION SYSTEMS

by

William G. Brown

DDC
JUL 6 1965

20040825052

April 1965

U.S. ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

Best Available Copy
ARCHIVE COPY

Best Available Copy

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

. Destroy this report when it is no longer needed. Do not return it to the originator.

AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC. DDC release to Clearinghouse for Federal Scientific and Technical Information is authorized.

AD 617 118

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, CFSTI
INPUT SECTION 410.11

LIMITATIONS IN REPRODUCTION QUALITY OF TECHNICAL ABSTRACT BULLETIN
DOCUMENTS, DEFENSE DOCUMENTATION CENTER (DDC)

☐ 1. AVAILABLE ONLY FOR REFERENCE USE AT DDC FIELD SERVICES.
COPY IS NOT AVAILABLE FOR PUBLIC SALE.

☒ 2. AVAILABLE COPY WILL NOT PERMIT FULLY LEGIBLE REPRODUCTION
REPRODUCTION WILL BE MADE IF REQUESTED BY USERS OF DDC.

☒ A. COPY IS AVAILABLE FOR PUBLIC SALE.

☐ B. COPY IS NOT AVAILABLE FOR PUBLIC SALE.

☐ 3. LIMITED NUMBER OF COPIES CONTAINING COLOR OTHER THAN BLACK
AND WHITE ARE AVAILABLE UNTIL STOCK IS EXHAUSTED. REPRODUC
WILL BE MADE IN BLACK AND WHITE ONLY.

TSL-121-2/65

DATE PROCESSED: 7-9-65

PROCESSOR: G. Lee

AD

DA-1L013001A91A
AMCMS Code: 5016.11.84400
HDL Proj 01201

TR-1295

A MULTIPLE TESTING OF THE ABC METHOD AND THE DEVELOPMENT OF A SECOND-GENERATION MODEL

PART 1
PRELIMINARY DISCUSSIONS OF METHODOLOGY

by
Berthold Altmann

SUPPLEMENT
COMPUTER PROGRAMS OF THE HDL INFORMATION SYSTEMS

by
William G. Brown

April 1965

U.S. ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

A Multiple Testing of the ABC Method and the Development of a Second-Generation Model

Part 1
Preliminary Discussions of Methodology

by

Berthold Altmann

Supplement
Computer Programs of the HDL Information System

by

William G. Brown

PREFACE

1. Major projects involving the development of a new storage and retrieval method, the organization of a fully automated library system, and the design and performance of a comprehensive test cannot be accomplished without dedicated cooperation if such projects are not endowed with separate funds and space and personnel allocations. The acknowledgement of invaluable contributions made by many constitutes, therefore, not so much a matter of form and politeness, but of understated truth and heartfelt gratitude.

2. We are extremely grateful to the members of other research agencies who spent many hours of their precious time joining in the evaluation of test questions and retrieval results: Messrs. T. Henton, D. Slater and H. Sullivan of the Advisory Group on Electron Devices, New York, who came to Washington for these tasks with the support of and at the behest of Mr. R. Dewitt, Office of the Deputy Director Research and Engineering, Department of Defense; Mr. C. Marsden, of the National Bureau of Standards; Mr. G.D. Goldstein and Lt. S.J. Mathis, Office of Naval Research, Navy; and last but not least the members of the Air Force Office of Scientific Research, Lt. Col. R.W. Conners, Dr. J.T. Ratchford, and Mr. M. Swerdlow, contacts with whom had been facilitated by Dr. H. Wooster.

3. Dr. Godfrey Knight, Jr. of the Cambridge Communication Corporation gave welcome assistance by having all test questions as well as several hundred documents classified according to the subject scheme of his organization. A special bibliography and copies of a great number of documents were the contributions of the Defense Documentation Center. Our research analysts, six professors of the School of Engineering and Applied Sciences, George Washington University, prepared the concepts under the pressure of deadlines. Their names are presented on the subsequent page. However, it is our duty to thank Dr. Louis dePian separately who in addition to the task of screening and standardizing the analyses prepared by his group has also contributed to the solutions of different difficult problems.

4. Only through the support of our Technical Director, B.M. Horton, could we obtain the cooperation of HDL scientists and engineers in the preparation, the performance, and the evaluation of our test. By listing their names below we do not do full justice to those whose diligent and careful evaluations and comments exceeded by far the general assignment.

5. We owe a special debt to Dr. W. Youden, NBS, for the review of our test program and especially for his suggestions on how to utilize our control group. Also Dr. B.M. Kurkjian and his team of statisticians at HDL helped in shaping the test procedure according to its true objectives. The personnel of the HDL library patiently and gracefully supported the establishment of the test collection

by moving their holdings to provide a special room despite their desperate plight for working and shelving space; and did join the ranks of the retrieval operators as a separate group. Mr. William G. Brown prepared or adjusted the required machine programs. An outline of his efforts leading to the automation of the entire library system is presented as a supplement. Although still very brief in its present form, it offers a basis for cooperation with those working on similar plans and objectives.

6. Miss Kathleen Rydlewicz completed with great dedication, unusual skill and deep understanding the numerous administrative tasks such as organizing the test collection, preparing and conducting the test, scheduling the operations and insuring the compliance with established deadlines by research analysts, computer programmers, key punch and machine operators, printers of forms and instructions and by all those who performed the test and evaluated the results.

7. Suggestions by Mr. Theodore B. Godfrey, Dr. W. Menden and Mr. H. Ogata, all three at HDL, and by Professor Thomas Wiggins of George Washington University have aided in improving the format.

HARRY DIAMOND LABORATORIES CONTRIBUTORS

Group 1A and B

Formulators of Questions and Retrieval Operators

1. C. Arsem
2. W. Boesman
3. P. Brody
4. J. Casey
5. T. Conway
6. E. Cox
7. J. Cronin
8. B. Dunmore
9. R. Garver
10. R. Goldfarb
11. F. Goodman
12. F. Gray
13. D. Green
14. P. Hawes
15. R. Hens
16. M. Hine
17. L. Hoffman
18. W. Isler
19. F. Liss
20. R. McCracken
21. O. Meyer
22. M. Miner
23. H. Morlock
24. E. Naess
25. G. North
26. H. Oelke
27. R. Paradis
28. R. Puttcamp
29. J. Ramer
30. E. Ramos
31. D. Readey
32. L. Salisbury
33. K. Sann
34. R. Sterns
35. R. Terrell
36. R. Tucker
37. F. Vrataric
38. D. Whittaker
39. W. Wiebach
40. L. Williams
41. K. Zastrow

Group 3

Library Personnel Retrieval Operators

1. E. Channel
2. E. Offutt
3. A. Porter
4. R. Ryan
5. J. Rosenberg
6. M. Taylor

Group II

Evaluators

1. C. Anstine
2. F. Balicki
3. W. Behrens
4. M. Brenner
5. J. Cooperman
6. D. Finger
7. T. Godfrey
8. G. Keehn
9. C. Klute
10. H. Kohler
11. I. Marcus
12. T. McCall
13. R. McCausland
14. B. Mulligan
15. M. Reddan
16. I. Rotkin
17. J. Scales
18. J. Tompkins
19. B. Willis

GEORGE WASHINGTON UNIVERSITY

Group 2

Research (information) Analysts

1. D. Anand
2. D. Baechler
3. N. Grisamore
4. N. Sloan
5. T. Wiggins

Standardizer

1. L. dePian

Table of Contents

		Page
Preface		I
Table of Contents		V
Abstract		VII
1	Introduction	1
2	The Test, A Preliminary Report	3
2.1	The Collection	4
2.2	Test Questions	6
2.3	Retrieval Operations	9
2.4	Evaluation Procedure	11
2.4.1	Preliminary Evaluation Procedure	15
2.4.2	Definitive Evaluation Procedure	16
2.4.3	Tabulation of Data	18
3	The First-Generation ABC Method in Operation	22
3.1	The Multidimensional Format	22
3.2	The Educational Capability	23
4	Characteristics of the Second-Generation ABC System	26
4.1	ABC Dictionary	26
4.2	Second Generation Card Catalogs	31
4.2.1	Parameters	31
4.2.2	Descriptive Information	33
5	Conclusion	35
5.1	Second-Generation ABC Method	35
5.2	Characteristic Aspects of Test Operation and Performance	37
APPENDIX A	The Cost Factor	39
B	Divisions of the Classified	41
C	(Memorandum) From: B.M. Horton, T.D. Cooperation in Test	43

			Page
Figure	1	Preparation of Test Collection and Retrieval Tools	44
	2	Preparation of Test Questions	45
	3	Guideline for Question Formulation	46
	4	Evaluation of Questions	47
	5	Retrieval Form	48
	6	Evaluation Procedures	49
	7	Evaluation of Test Results	50
	8	Retrieval Loops	51
	9	Statistical Analysis(form sheet for questions)	52
	10	Statistical Analysis (form for retrieval operators)	53
	11	Statistical Analysis (form for overall summary)	54
	12	Plot of First-Generation Retrieval Process	55
	13	First-Level Proliferation of Inquiry by Interrelationship of Concepts	56
	14a - 14e	Plot of First-Generation Retrieval Process	57-61
	15	Second-Generation Subject Organization	62-64
	15a	Organization of Concepts in the Second-Generation Dictionary	65
	16	Retrieval Using "ABC" and "Coordinate Indexing"	66
Chart	I	Test Operations by Groups	67
	II	First-Generation ABC Dictionary	68
	III	Second-Generation ABC Dictionary	69
	IV	Sample of Parameters (First-Generation)	70
Supplement:		Computer Programs of the HDL Information System	71
	I	Present System	72
		A Cataloging System	72
		B ABC Dictionary Update System	74
	II	Planned System	75
Figure	A	Over-all Planned System	79
	B	Present Cataloging System	80
	C	ABC-Dict Update System	81
	D	Library Cataloging Worksheet	82
	E	Library Cataloging Worksheet	83

Abstract

The first-generation ABC storage and retrieval method, an HDL-developed method that utilizes appropriate standardized English-language statements processed and printed by a KWIC-type computer program, was subjected to a performance test at the request of DOD. This is the first descriptive part of the test report ; a statistical analysis is in preparation.

The four groups responsible for the plans, the performance of the test, the evaluation of the results, and the statistical analysis of the tabulated data were kept separate from each other. With the exception of the designers of the test and the analysts (University professors) all operators were volunteers. The designers prepared the plans and instructions, established the test collection of 3650 magazine articles and technical reports on solid state devices, and supervised the technical operations. Two groups comprising about 40 scientists and engineers of this installation formulated two sets of test questions (100 from a random selection of the articles and reports and 36 from a general knowledge of the subjects covered by the collection), performed the retrieval runs, and pre-evaluated their own results.

A group of 30 senior scientists and engineers (including subject specialists of other agencies) evaluated and standardized all questions, and later evaluated all compiled data.

Three methods, two variations of the first-generation ABC method together with a KWIC-title list, were tested. A control group was established to discern the bias possibly introduced by the use of the test operators' own questions and the sequence of the three test runs. In addition to the 40 scientists and engineers (divided into two groups) mentioned above, the six research analysts (University professors) and six HDL librarians searched the 136 questions by all three methods. Therefore, 1632 retrieval sheets containing more than 6000 documents were obtained.

The "relevance" and "recall ratios" will be calculated and used to evaluate the retrieval method. For this reason, criteria were established for the determination of "relevance", and procedures (retrieval loops) were introduced for the identification of the relevant documents missed during the test. To facilitate the evaluation, each evaluator was made responsible for subject divisions covering about 400 titles of the systematically arranged test collection.

Two retrieval runs are described in detail to point out the multidimensional characteristics and the educational capability of the method.

A cost account of the test and a brief description of the computer programs for the entire system, several of which facilitated the preparation and reproduction of the various catalogs, ABC dictionaries, and control tools, are appended.

The test was performed with the primary objective to spot deficiencies and to develop the second-generation ABC model. The results are briefly reported. The improved model is characterized by descriptors of unlimited length, the introduction of facets or microschedules which produce logical organization of documents under important keywords, and a decrease in the number of verbalized concepts (or statements). Type of document, level-of-difficulty descriptions and operating parameters of equipments (a feature of the ABC method) are transferred to card catalogs.

1. INTRODUCTION

In a previous report, we presented an outline of a completely automated library system.[1] The data first inserted into the library system (that is, during acquisition or cataloging operations) are in such a form and of such a completeness that there is no need for any retyping or secondary reproduction to issue accession lists, bibliographies, and indexes; to prepare book catalogs, catalog cards; and all administrative records required to renew periodical subscriptions, to communicate with the book binders; to disseminate and circulate (charge and discharge) books, magazines, and documents, and to recall overdue loans. The system will also automatically prepare purchase orders for items that have been selected. What is being applied consistently is a very simple principle: that no line once written to describe an item already in or desired for the collection will be typed or punched a second time for any other purpose or operation.

Although we have made considerable progress in writing and testing the machine programs for this system (See Supplement), this report will deal only with one major segment of that system: the method of evaluating and analyzing items in the collection and the procedures developed for organizing, storing, and retrieving them. This system has been called the ABC (Approach-by-Concept) method. It has been designed and operated in HDL for the efficient and pertinent dissemination and recall of very specific technical information.

We can here emphasize only a few of the characteristics of the method. It uses the natural language of the scientist and engineer to index and retrieve information, as the most economical means of providing meaningful access to the collection. A key to this method is an easily comprehendible but very specific index (dictionary) of the collection. The collection, needless to say, must be selected, analyzed, and indexed with equal care.

In obtaining a response to specific problems or questions, there is no need for a confrontation between the scientist and the documentalist (a generalist who tries to interpret the query of a specialist in terms of a standardized vocabulary). Nor is there a need for the documentalist to approach the big black box (a computer or any other memory device) with a program and a prayer that the magic words he has selected will elicit a cogent reply, along with not too many others, to his client's request. These two steps, common in most retrieval operations, are eliminated because they are the primary causes of misunderstandings, errors and confusion, both in indexing and retrieving. Instead, the investigator himself consults the dictionary of concepts. The lines in our dictionary are

---

[1] Altmann, B. "The Medium-Size Information Service; Its Automation for Retrieval," TR-1192, Harry Diamond Laboratories, 30 Dec 1963.

henceforth referred to as "concepts" although we recognize that they are verbal expressions. The term has been chosen to distinguish concise, meaningful and self-explanatory statements of control from other subject approaches such as subject headings, uniterms, descriptors, annotation, etc.

While using the ABC method, the investigator is forced to compare his original, often hazy formulation of a problem with the standardized descriptions of available information. It is therefore anticipated that he will increase his knowledge and understanding of the problem with which he is confronted by refining and restating his question with greater precision as he makes his way through this multidimensional system. He will recognize the full complexity of his problem as he views it (a) in the context of similar, parallel, or slightly different efforts and achievement, (b) in relationship to the specific and general aspects of his problem, and (c) in the relationship to the directly and indirectly associated scientific and technical fields.

During the past year, the study and further development of the ABC method concentrated on three different tasks. The first task was to design and perform an objective test of the system and to determine the validity of major assumptions and claims that had been made regarding the system. A test system was built; a test was conducted; and the evaluation and statistical analysis of test data has begun.

Members of outside agencies were invited to join in the development of adequate and fair testing procedures, participate in the actual retrieval operations, and help in evaluating results. Because of its classified nature, it was not appropriate to use the installation's collection and ABC dictionary in the test. Consequently, a separate test collection was established and special dictionaries for this collection were prepared for the test.

The second task was a critical assessment of the first-generation ABC method, based on several typical retrieval operations, to substantiate previous claims with respect to its multidimensional characteristics and its educational capability.

The third task covered in this report deals with a second-generation ABC retrieval method. This system is based on experience gained in the test; on the analysis of difficulties scientists, subject specialists, and librarians encountered as they used the first-generation system; and on a previously conceived ideal storage and retrieval system. This second-generation system superimposes small logical classification schemes or microschedules upon the existing format to streamline the dictionary and facilitate retrieval.

In Appendix A, we respond to the request of reviewers of the first report by supplying the cost figures compiled during the test, which may help in calculating the expenditures required for introducing and applying this retrieval program in a particular agency.

2. THE TEST: A PRELIMINARY REPORT

The test of the ABC method was conducted at the suggestion of Mr. Walter Carlson, Director of Technical Information, Office of the Director of Defense Research and Engineering, Department of Defense. The work was performed within the budget and the personnel ceilings of the Technical Information Office in HDL. This will explain the unavoidable utilization of sometimes reluctant "volunteers" recruited in our laboratories and, in particular, the delays in scheduled operations, and possibly, for some unevenness in performance and test results. Far outweighing these disadvantages, however, were great advantages: (a) testing of the ABC method by persons with negative attitudes and a very limited knowledge of the system lends some credence to such favorable evaluations as may result; (b) personnel most interested and knowledgeable, connected with research operations of this type in DOD, AF, Navy, and NBS, participated in the planning and evaluation phases and provided the necessary objectivity and expertise; and (c) it was possible to place most functions under strict controls, and to identify variables, so that unbiased evaluations could be obtained.

A requirement imposed on the test, which might have limited its scope to some extent, was adherence to certain procedures and methods that personnel at the College of Aeronautics, Cranfield, England, had introduced when they, under contract to the National Science Foundation, compared the relative efficiency of four different indexing systems.[2] Only by following the general outline of the British test could HDL produce results amenable to comparative analysis. To circumvent certain disadvantages in this approach, however, we took the liberty of additional control test runs and of making changes and adjustments to duplicate situations more common to our experience.

The basic requirements of the test were obvious: (a) we had to assemble a representative test collection, cataloged, processed, and organized in accordance with the principles of the ABC method; (b) we had to generate a set of pertinent questions which the collection could or perhaps could not answer; and (c) we had to establish standard operating procedures and controls to insure that the test was valid, and the standards for evaluating its results fair.

---

[2] Cleverdon, Cyril W., ASLIB Cranfield Research Project - "Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems." Oct 1962. Cp. also Aitchison, Jean and Cyril Cleverdon, ASLIB Cranfield Research Project - "A Report on a Test of the Index of Metallurgical Literature of Western Reserve University." College of Aeronautics, Cranfield, England, Oct. 1963.

2.1 Test Collection

Because it was recognized that the collection was one of the critical factors demanding special attention during the preparatory phase, we first consulted our statisticians, who recommended the customary formula: the size of the test collection should be at least ten times larger than the number of papers we expected to retrieve during the various test runs. We intended to formulate and use at least 100 questions and to obtain an average of three to four answers so that an adequate basis for determining relevance and recall ratios would be provided. Therefore, we arrived at a required figure of 3500 titles; the actual size of the collection later turned out to be 3650, consisting entirely of journal articles and technical reports.

We provided for a reasonable recall potential by requiring a collection in depth; that is, we narrowed (and thereby deepened) the scope of the subject area to solid state devices, circuits, and applications. The subject area encompassed a variety of principles and products, such as dielectric, magnetic, conductive, and photoelectric devices, their uses as amplifiers, oscillators, switches, and pulse generators; and their use in communication, control, computing systems, test instruments, etc. The subject area was sufficiently small to permit comprehensive coverage with 3650 items. Therefore, even though small, the collection provided for a satisfactory number of pertinent replies to a given question.

Our selection of this particular subject area was prompted by a number of other reasons: (a) it provided useful, valuable, and timely information (published after 1959) for personnel of our installation as well as for the information analysts (George Washington University School of Engineering and Applied Science) used in the test; (b) it made cooperation in the test project more attractive to the "volunteer" retrieval operators in HDL, since the subject area was applicable in their work; (c) it expedited the critical analyses for indexing because the information analysts were familiar with the subject; and (d) it facilitated the establishment of an additional tool by introducing a conventional subject card with multiple entries and abstracts, which very effectively enabled evaluators to check on the completeness of the retrieval and to determine the recall ratio.

This auxiliary control catalog was based on catalog cards and abstracts supplied by the Defense Documentation Center (DDC) for technical reports and similar cards for journal articles supplied by the Cambridge Communication Corporation (CCC). This card catalog was organized according to a detailed subject classification scheme prepared by CCC. The condensed classification scheme is given in Appendix B.

The mechanics of assembling the test collection and retrieval evaluation tools are shown in Figure 1. The individual items for the test collection were selected from various bibliographies particularly from the Solid State Abstracts published by CCC and a special bibliography issued by DDC. Reprints of articles were obtained from various libraries, and reports were furnished by DDC.

The selected papers and reports were sent to the information analysts for evaluation and formulation of ABC concepts. The concepts were then standardized and transferred to punched cards. In a parallel effort, the items were descriptively cataloged, and the descriptive titles key punched. Both types of information, the concepts and the titles, were then combined and transcribed by an IBM 1410 computer on a catalog file tape following a sort and merge program.

Using the catalog tape file, a 7090 computer produced eight catalogs: a and b) two ABC dictionaries, both listing (following a KWIC program) the concepts as well as the term-letter code combinations[3] (under which the corresponding titles are arranged in the ABC card catalog) but differing with respect to the number of alphabetized keywords; c) an ABC catalog with individual titles and accession numbers filed under the appropriate alphabetically arranged asterisk term and code; d and e) an alphabetical list of letter codes with the concepts they signified and a corresponding card file; f) a card file of accession numbers giving the titles represented; g) a KWIC title list with significant words rotated and alphabetized; h) a file of reports listed numerically under the AD numbers.

In contrast to the "short" dictionary, all nouns, adjectives, verbs, and numerals were treated as key words in the "long" dictionary. (The selection of keywords in the former dictionary was made by the information analysts during the standardization process.) Consequently, the short dictionary was only half as long. The KWIC list of titles furnished a control access to the collection. The other tools unessential to the ABC system were generated to aid in the evaluation of test results.

In a separate effort (and not by a machine process) the abstract cards prepared and published by CCC and DDC were organized into a subject card catalog.

---

[3] For samples see Chart II.

2.2 Test Questions

After the test collection had been organized and after both the ABC and alternative approaches to the collection were established, we faced the formidable task of formulating the appropriate questions for use in the test. If a realistic situation was to be simulated, these questions had to be addressed not to the wording of the titles but to the contents of the test collection, and they had to be meaningful in being concerned with the actual problems and interests of our installation. Moreover, the majority of the individual retrieval operations had to yield positive results if the capability of the method was to be demonstrated. Also, it was desirable that the entire test collection participate to some extent in the operation. Realizing this, the procedure outlined in Figure 2 was adopted.

The computer was used to select at random, 400 titles from the systematically organized catalog of the collection. This was printed as a list providing no more information than (a) the subject classifications that had been supplied by CCC; (b) for journal articles, the title of the periodical, volume number, and page number; and (c) only the DDC accession numbers for the reports. From this list, each of 41 scientists and engineers, assigned to the task by request of the Technical Director (Appendix C), selected subject areas consistent with his interests from which to derive test questions. Despite the freedom of choice given to the individuals, it turned out that all the subject areas represented in the collection and included in the random list were well covered and quite evenly distributed.

The members of this group (identified hence as Group 1)[4] then selected documents of interest insofar as such interest could be determined from the information available at that point. The titles were not known to them until the item was supplied. They then examined the text, and if they elected to use it, formulated a test question that would be answered by its content. They had been reminded at the start of this operation that because of economic limitations, the information analysts could not describe or index each incidental statement or each casual remark made by the authors but could pay attention to and process only such substantial aspects as the objectives and methods of the investigation; the devices, components, and materials discussed; their properties, processes, and instruments; the pertinent parameters; and especially, new results. This guidance was given in oral and written form to insure that only substantial questions would be formulated.

The product of each individual effort was a worded question which together with the accession number of the basic document, and the worker's name and laboratory was entered on a specially prepared form,

---

[4] For the various groups, their designations and operations, see Chart I.

which with the instructions given is illustrated in Figure 3. We obtained about 225 questions in this manner. However, a method basing all questions on the contents of the collection hardly produces a true-to-life situation.

To provide a degree of realism, we submitted an outline of the subject categories covered by the collection (Appendix B) to scientists and engineers who, because of their scholarly attitudes, maintained year-round contacts with our office, and requested them to formulate additional questions based merely on their general knowledge of the subject areas of the collection and on their own experience in the laboratory. In response to this request, they provided 36 questions without reference to any particular document.

We had anticipated that several contributors would choose identical papers or submit questions similar in formulation or substance. For these, and other obvious reasons, it was necessary that the questions, prior to use in actual retrieval operations, be evaluated, combined, and edited to eliminate deficiencies and redundancies. Therefore, the questions were transferred to an "evaluation of question" form (Figure 4). To preclude bias in this process, these evaluative and editorial responsibilities were assigned to a group (henceforth identified as Group II) of senior scientists and engineers. Because we charged them also with additional tasks of making final decisions and of exercising controls, we not only kept them completely separate from members of Group I, but also supplemented their roster with personnel working for the Department of Defense, the Air Force, Navy, and the National Bureau of Standards. Many joining this group stipulated that the time required for these tasks could not exceed an average of 6 to 8 hours. To comply with this demand, we increased the membership to about 30.

To enable the members of Group II to perform their assigned tasks within the limited time allotted, the following preparations were made: (a) The 265 questions were organized by the subject scheme used in the subject card catalog. (b) The systematically arranged questions were divided into 10 major sets of 25 to 30 questions, each set corresponding in theory at least to one tenth of the test collection. (c) Each set was then assigned to an editorial team of three, having an interest in that particular subject area. (d) The questions were then divided equally so that each member processed about eight questions. The evaluator compared the text of the suggested question with the contents of the original paper, and decided to drop, combine, rephrase, or approve the submitted questions. He had the prerogative of discussing complicated problems with members of his team or representatives of the Technical Information Office, but the decision was his alone. He could, if he so desired, anticipate results of the test and locate the titles of papers applicable to the questions and in general was in a position to familiarize himself with that portion of the collection allotted to him. As a rule, however, the results were evaluated at the completion of the test runs, at which time complete familiarity with the assigned portion of the collection became necessary. (e) The

evaluators entered their decision in the space provided on the form with the question (Figure 4), listing their finally approved version of the question.

The 36 questions formulated without benefit of a particular paper were standardized only with respect to clarity, pertinency and adequacy.

As a result of this process, 100 questions formulated by Group 1 were approved in addition to the 36 questions formulated without benefit of a particular paper. All questions were transferred to retrieval forms (Figure 5). Twelve retrieval forms were completed for each question, to provide multiple retrievals, using various tools in various sequencies, and by various groups.

2.3 Retrieval Operations

Four different groups (Figure 2) performed the retrieval operations using the identical set of 136 questions. These groups included (1) two subgroups formed from those in Group I whose questions had been accepted (Group 1A and 1B), (2) the information analysts (Group 2), who had evaluated the documents and had formulated the concepts for the dictionary, and (3) personnel in the HDL Technical Information Office (Group 3), who performed general reference services. Members of the latter two groups were homogeneous and had no preference regarding subject specialty. Each was asked to retrieve documents using tools in what was called the normal sequence: (1) the short ABC dictionary, (2) the KWIC list of titles, and (3) the long ABC dictionary. Members of the two subgroups (1A and 1B) were assigned questions in subject areas in which they had prepared questions both that they had formulated themselves and that their counterpart in the other subgroup had prepared. They were asked to process first their own questions, and then, their counterpart's.

The subgroup 1A was asked to use the tools in normal sequence in processing the questions. To provide for additional evaluations, subgroup B was asked to use the KWIC title list first and the short ABC dictionary second.

To enforce the prescribed sequential use of the tools, the individual operator received at one time all his questions (each recorded on a separate form) for testing one of the three approaches, let's assume the "short" ABC dictionary. A member of Group 1A with an average assignment of three questions of his own, and three of his counterpart's formulation was to turn in the answers to all his questions before he could receive a duplicate set for processing the six questions with the second tool, the KWIC title list. The freely styled questions are not considered at this time. We established this procedure to create an interval of at least one day between the two retrieval operations for the same question; and made the assumption that because of the lapse of time the retriever had forgotten the alphabetical codes and accession numbers he had recorded during the first run, and that his previous experience would not influence the results of the second run.

Although space was provided on the retrieval form (Figure 5) for the basic document, this information was withheld from all groups in a major deviation from the Cranfield test. After the basic document had been processed to provide a pertinent question (a process we had considered artificial and therefore accepted only reluctantly), a measure of realism was introduced in that the retrieval was conducted freely without knowledge of the "answer." It was felt that the retrieval of the basic paper could not and did not signify the success or the end of the operation, and withholding the accession number insured this. Later, retrieval of the basic document is to be rated of major importance only if it provided an excellent answer, and of little consequence if other items retrieved furnished better or more substantial

replies. It was to constitute a non-critical factor in the final evaluation.

The retrieval itself consisted merely of filling in the start and stopping time, and identification of concept (the asterisk-term and code) used and accession number of applicable titles that were found under the concepts in the ABC card catalog. If the tool was the KWIC list, only the accession numbers were entered. The four groups filled out a total of 167 retrieval forms.

BEST AVAILABLE COPY

2.4 Evaluation Procedure

In the preceding chapters, we described a sequence of procedures and operations intended to produce the test data. Before we continue with the description of the subsequent phase, the evaluation of the data, we pause for a brief account of the principles underlying our procedure and our belief that we will approach our envisaged goal with the types of data we have assembled.

What we attempted to accomplish was a test of the ABC storage and retrieval method and an evaluation of its efficiency in terms which permit comparison with similar systems. As in every operation of this type, we must create and apply appropriate yardsticks for taking quantitative measurements of the performance. Because we follow the example of the Cranfield program, we will establish two ratios: (a) the relevance ratio, a measure of the system's utility expressed as the percentage of useful or relevant items recovered in a given (or average) test run; and (b) the recall ratio, a measure of the system's efficiency in terms of its capability to identify or recall pertinent papers embodied in the collection in response to a particular (or average) question.

In theory, these are reasonable measurements of the service of an information office provided that we are in a position to rate an average performance in terms of user satisfaction and capability of locating all appropriate titles in the storage system. In practice, however, this cannot be easily accomplished. With respect to reader satisfaction, we know only too well that despite a common interest in the same subject matter preference with respect to an individual paper may differ widely between scientist and engineer, physicist and chemist, junior scientist and expert, generalist and specialist; and because his knowledge and understanding have grown, the same reader may reject today what he cherished a year ago.

The factor of human fallibility poses another obstacle to consistent evaluations and objective comparisons of different storage and retrieval systems. Librarians may have acquired materials that are not appropriate, analysts may have prepared descriptions that are ill-suited, investigators may have been ineffective in describing their true requirements, and retrieval operators may have missed the correct approaches to the pertinent information contained in the system.

How can we determine the relative merits of two systems, if, not only the systems, but also the collection, the analysts, their education and experience, the types of data, the methods of obtaining them, and the conditions under which it is done, differ.

In fact, the views on the validity or the mere usefulness of tests vary drastically. One can argue that the difficulties and differences encountered in testing only reflect the variables present in all operational systems and should therefore be dismissed as a cause for major concern. This would render the comparison of systems

most difficult, if not impossible. Conversely, one can consider the variables so disturbing and the possible results so unreliable, that a test appears to be no more than an exercise in futility.

When we are exposed to these arguments, we must remember that storage and retrieval systems are not the only ones greatly influenced by environmental factors and by the quality of human performance. Nevertheless, man-operated systems, weapons, and equipment have not only been tested, but their performance compared with that of rival systems and equipment.

Test design and test procedures have been greatly refined. When the enormous cost of modern, complex military systems made it prohibitive to use a large number of units or field tests, the engineers designed bread-board models of the various subsystems, devices and components, studied them with greatest intensity, subjected them to tests simulating the operational environment, and redesigned and rebuilt them until they exhibited the required performance reliability.

To a great extent, we followed the same procedure in the test of the ABC method. When the collection had been assembled and the catalogs and tools for retrieval and control prepared, we had no intention of defending "our investment" and proving the efficiency of the ABC method. For us the entire operation was an experiment rather than a test, it was the welcome opportunity to subject to vigorous simulation the bread-board model we had built, to determine its deficiencies, to analyze the faults and failures of the total system as well as of its individual components, to seek new solutions and especially to redesign or adjust the retrieval methods, to reduce the human mistakes and render the retrieval operations of the information-seeking scientist and engineer (for whose immediate use they were developed) simpler, faster and cheaper.

In a subsequent chapter, we will present two samples of our test method to demonstrate the multidimensional and educational characteristics of the ABC method. At the time that these tests were performed it was their primary purpose to identify trouble areas and assist in the redesign of the system. This critical attitude prevailed throughout the entire test period. It was a constructive operation "against" and not "in favor" of the system.

There is a particular cause for this personal detachment. As the test progressed at a slow pace, we discovered deficiencies, analyzed their causes, searched for appropriate solutions, and redesigned and replaced the faulty parts. However, personnel spaces and other support necessary to retrofit the retrieval system of the test collection were not available whenever possible improvements became apparent. We were only able to continue and complete the test of the model in its original form. When we passed the half way mark of the test, most of the procedures should have been changed, and what we were still subjecting to our test procedures was a de-facto obsolete model.

It was obvious that we became increasingly preoccupied with our work on the second-generation ABC method, and that we had no longer a personal stake or a major interest in the outcome of the current test.

To insure the reliability of the data obtained during the official test runs, we organized the four groups of retrievers to resemble in composition and proportion the profile of actual users. Because the system was designed primarily for the scientists and engineer at the work bench, they constituted 77.3 percent of the operators and each (as in reality) was responsible for a relatively small and specific subject area (that is for about 3 percent of the entire questions). The George Washington University professors were assumed to represent the senior scientists and engineers; they numbered only six (or 11.35 percent of the total), but covered 22.7 percent of the total number of questions per person and in numerical retrieval output, the HDL librarians (4 reference librarians and 2 catalogers) equalled the group from the George Washington University.

In order to obtain reliable test results, we had to limit the human error factor, identify the bias eventually introduced by the testing procedure, and "objectivize" the evaluations. Although we had assumed the task of determining the capability of the system as such, we remained fully aware of the extensive intrusion of the human element in the preparation and in the conduct of the test. This necessitated the formulation of standards or yardsticks for discovering and discounting the questions that were poorly phrased (in relation to the basic paper), the concepts that were adequately prepared by the analysts or incorrectly selected and applied by the retrieval operators, and the evaluations that had been influenced by the bias of the evaluators. The methods used to accomplish these ends were no different than those usually applied by professional test engineers: critical analysis, generation of multiple test data (repetitious test runs and evaluations), and the introduction of control groups and stringent controls.

The disquieting connotation of relativity and subjectivity generally inherent in the term "relevance" was removed because we did not consider relevance with respect to a person or group of persons but rather by comparison with the contents of a document or the substance of a question. However, we were not satisfied with these still elementary methods, and preferred to determine "relevance" more objectively. We analyzed the conceptual substance of the question, attached relative weights to the various conceptual components and their combinations, and used the resulting scales as a measuring stick.

In the process of these evaluations, we came to a major deficiency of the first-generation method. The concepts identified publications by form (e.g. bibliographies, collections, symposia, etc.), by purpose, treatment, and application (that is methods of origin such as analysis, test, design, development, experiment, etc.) and by level

of difficulty. As a result of this observation, the second-generation model (to be discussed in Chapter 4) will permit retrieval of documents not only by subject approach, but also by personal preference and will therefore improve the relevance and recall ratios even though subjective standards may be applied to the worth of the documents.

The test questions were processed eight times according to the ABC method. Because the test data taken for each question were tabulated on a separate form, we obtained meaningful evaluations in a very short time.

In the eight test runs when all of the operators failed to produce the basic document, it can be assumed that the analyst did not provide an appropriate conceptual approach. For verification of the assumption the evaluator should first turn to the table in which concepts as well as questions are displayed with the titles of the documents from which they were derived. This table has already been prepared; together with its analysis it will be published in the report of the statistical results.

If, on the other hand, the basic document was located by one or more retrieval operations, the instances of failure may be traced to an oversight by the operator, to the brief presentation of the concept in the short-form dictionary, or to the formulation of the question.

The tabulation of the results by operator will facilitate analyses of performance by groups with respect to retrieval tool, time and success or failure of the retrieval if the sample should prove to be of sufficient size; it may permit the identification of certain individual qualities (educational and professional background, reading habits, scholarly attitudes, achievements, etc.) which may predict the success of an analyst or retrieval operator.

The eight different test runs for each question may help to determine whether the causes for failure may be traced to the human element or to the system. The multiple evaluations of the test results by Groups 1 and 2 and the evaluators in Group II will provide information about the range of personal opinions (e.g. with respect to relevance) and will permit a confrontation of subjective and objective evaluation.

The absolute or relative value of multiple data gained during the test of a retrieval system cannot be determined before the completion of the statistical analyses. If it should be established that repetitive tests by representative groups of operators provide a more reliable basis for evaluations, the repetitive tests should be performed with identical collections to reduce the variables by one predominant element.

For the test of our overall library system and for the economical production of all our different retrieval tools, the catalog of our test collection was transferred to magnetic tape. Copies of our catalog can, therefore, be provided cheaply and rapidly for use in testing other systems under comparable conditions. In addition we preserved copies of a number of articles and of all reports that formed the test collection. In this connection, we would like to suggest that catalogs of test collections should always be available for testing of other systems.

Control factors and control groups have been used to insure realistic conditions, reliability and accuracy of data and consistency of testing procedures. Although we will in these paragraphs limit our brief summary to controls exercised with respect to evaluation, we cannot avoid repeating information we have previously mentioned.

The three retrieval loops we developed are control "mechanisms" to assist the evaluator in discharging his heavy responsibility which is the recall from the collection of all titles of relevant papers which the operators had failed to retrieve.

The most important control was exercised by the members of Group II (composed of senior scientists from HDL and other agencies and kept entirely apart from all the other operational groups). It was their task to evaluate and to approve the questions and to control the evaluations of the retrieval data in terms of the measurements of completeness and relevance.

Of the retrieval groups, those in Group 1 (the HDL scientists and engineers) and in Group 2 (the information analysts) had a personal as well as professional interest in the outcome of the test. Therefore, they were allowed to make only preliminary evaluations of their own results. The evaluators (Group II), being largely unbiased observers, were assigned the task of making the detailed final evaluations.

The complete results are presently being tabulated, and statistical analyses are being prepared. However, the nature of the results may be inferred from the evaluation procedures which follow.

2.4.1 Preliminary Evaluation Procedure

Preliminary evaluations were performed by Groups 1 and 2 according to the procedures indicated in Figure 6. Group 3 was omitted from this task.

For the purpose of preliminary evaluation, the evaluation sheet shown in Figure 7 was used. Entered on the form were the name of the retriever, his group or team designation, the question, the accession number of the basic document if one existed, and the

accession numbers of all documents the particular retriever located during the three runs in response to the question. The retriever then evaluated the contents of the retrieved documents, compared them with the basic document, and checked the appropriate column to indicate their value: equal (=), better (+), inferior (-), or not applicable (0). Questions that were not based on specific documents were graded on the quality of the answer. The same form was used for the preliminary evaluation by the retrieval operator as well as the final evaluation by the assigned member of Group II. In the latter case, the accession numbers of all the documents identified in response to the question (during the 12 test runs) were listed in the left column of the form. In addition, a separate column was reserved for the evaluator in Group II to assess pertinent documents which had been missed by the retrievers.

2.4.2 Definitive Evaluation Procedure

The twelve retrieval forms and two evaluation sheets for each question were returned to the same person in Group II who had standardized and approved the question. This group evaluated the results as follows (Figure 6).

Using the CCC organized abstract card catalog, the pertinent titles not retrieved during the test runs were determined and graded in the second column of the evaluation form (Figure 6). In order to accomplish this somewhat impossible task, the ten teams of Group II had to acquaint themselves thoroughly with that section of the collection assigned to them. This provided the basic data necessary to determine the recall ratio, that is, the efficiency with which the system retrieved all relevant information in the test collection.

Since it was desirable to determine the cause of the wrong choices or the wrong failures to choose documents, an evaluation of all concepts used in a given run was made (Figure 6). To make this evaluation as objective as possible, the following procedure was followed. The essential elements signifying the contents of a question were identified, and the combination of them that would have appeared in appropriate concepts were theorized. The theorized combinations were individually graded as +, =, -, or 0, and the concepts used in retrieving were graded by comparing them with the theorized combinations.

This provided a basis for judging the use of a concept as proper (+,=,-) or as improper (0) and therefore assignable as an operator error.

To further illustrate this process, the specific analysis is given regarding the question: generation of high frequency energy in semiconductors. The persons who evaluated the concept determined first that the different substantive elements were: 1) high frequency generation; 2) generation takes place within semiconductor materials; 3) generation is accomplished by or with semiconductor materials; and 4) high frequency energy. In this process, the elements were merely identified and enumerated casually.

In grading or ranking the elements, the following weights were assigned:

+ on the combination of 1, 2, 3
= on the combinations 1, 3, 4, or 1, 2
- on the combination of 1, 3
0 on all other combinations and on the single elements

Only after the establishment of these weights or standards did the evaluator rate the various concepts that the retriever had used in the different runs.

In a number of instances, the same combinations of elements were assigned two different grades, especially = and -, and the quality of the papers retrieved was used to assign the relative value of the concept. Inasmuch as the gray area doubt (-) was introduced, it can be assumed with a high degree of certainty that the concepts graded with (0) had been incorrectly selected, and such a negative result suggests an operator error that should not be held against the system.

Although a subject classified card catalog with abstracts had been made available, the evaluator encountered a difficult task in locating all titles in the collection related to the formulated questions. To ensure as thorough a job as possible, additional tools were provided as shown in Figure 8, and the following procedure was indicated.

The evaluator was to first turn to the ABC dictionary to find additional approaches missed by the members of the three retrieval groups, and if successful, was to follow the standard retrieval method; i.e., he noted the respective asterisk-term and code combination and checked the pertinent titles in the ABC card catalog (I). The tools and catalogs are identified by the Roman numerals (Figure 8). This first step could well yield new useful information. If additional concepts were recorded at the bottom of the card in the form of asterisk-term code combinations, he was to determine the complete text of the concept using a list of concepts with their codes arranged in alphabetical order (II). If it appeared to be applicable and was missed in the test retrieval runs, he was to return to the card catalog to complete this loop. He was to continue in this manner as long as new asterisk-term and code combinations turned up in the card catalog and until new titles could no longer be located.

He then utilized the second loop. From a very comprehensive alphabetical index of subject headings prepared by CCC, the evaluator was guided to the respective subdivisions of the subject card catalog (II) where all related information was combined. This catalog consisted of multiple title cards with full-length abstracts and numerous cross-references arranged according to a logical scheme. CCC had created this system and had also given our project most valuable

assistance by organizing the DDC reports in the test collection as well as our test questions in accordance with this scheme.

When the evaluators discovered a title in the indexed subdivisions that seemed pertinent, they could frequently determine its relevance by the abstract alone. They then noted the accession number, which was used to identify the combination of asterisk-term and code under which it was filed in the ABC card catalog. This was accomplished through catalog (IV). Following this, they examined the ABC card catalog (I) for the location of secondary concepts and exhausted this loop as before. These steps were repeated as often as new subdivisions and new titles could be found using the sequence of subject card catalog, new asterisk-term and code combinations, and the ABC card file.

Finally, the evaluator entered the third loop, screening the KWIC title list (V) under all possible pertinent and significant terms. If he was successful in finding an item omitted, he was to record the accession number printed with each rotated title. As in the preceding loop, this number would lead first to the title catalog (IV) with its asterisk-term and code symbols and second through the code to the ABC card catalog (I) where he could find secondary concepts to close the loop.

Whenever additional titles were found by any one of the three described recovery methods, the contents of the recalled materials as well as their underlying concepts were to be rated according to the same standards used to evaluate the original findings of the retrieval operators.

Although this rather elaborate scheme was provided and followed in a number of cases to obtain all applicable documents on a given question from the collection, in practice, the method was found to be much too repetitive. Therefore, the evaluators primarily relied on the second retrieval loop.

2.4.3 Tabulation of Data

The statistical data and their analysis will be presented in great detail in a subsequent study currently under preparation. In this preliminary discussion of our testing procedure it will nevertheless contribute to a better understanding of our efforts if we describe briefly the test data being processed for final analysis by the statisticians. On a summary tabulation, there are three major considerations:

1. The retrieval of documents for 136 questions by 4 different groups, each applying three different tools or methods had yielded 1632 retrieval sheets with a minimum average of three responses on each; in other words about 5000 data had to be organized for meaningful analysis;

2. data recalled had been evaluated twice, by the retrieval operators[5] as well as by Group II, for quality and pertinency with respect to the basic document used for the formulation of the question. If a document was not used to formulate the question, then the evaluation was with respect to the objectivity and scope of the question itself; and finally the concepts chosen by the retrieval operators had also been rated for the purpose of making allowance for human errors;[6] and

3. test had as its main objective the evaluation of the system in its realistic environment, so that this natural environment and its tolerance of human error had to be determined.

The test data can, therefore, be logically presented on three types of forms: the first (Figure 9) providing all the responses given to one particular question; the second (Figure 10) furnishing the evaluated results obtained by each individual retrieval operator using each particular tool; the third (Figure 11) summarizing the results by groups of operators.

The first form listed the question and identified (whenever applicable) the basic document and its concept. The test data were tabulated separately for each of the four retrieval groups by giving the total (N) of the documents recalled or concepts used in every run and the evaluations by retrieval operators[7] and by subject specialists (Group II).

In the retrievals by Group IA and IB, there were two possibilities of working with either their own or their counterpart's questions. This is provided by the "own ques." or "other's" lines. In each case, the concepts are always those of the particular retrieval operator. Thus, for each retrieval group and each retrieval run the concepts actually selected were evaluated.[8] The results are entered on the third line of each block under the heading "Subject Specialist."

---

[5] The librarians (Group 3) had been excused from this task.

[6] This method was applied also to documents retrieved by the KWIC-Title list.

[7] The librarians (Group 3) were excused.

[8] For the 36 free questions the documents retrieved were re-evaluated by a member of Group 2, the information analysts. His evaluations (figures in parentheses) were recorded: The concept evaluation is on the second line, the document evaluation on the third line.

The compilation of this form permits question-by-question comparisons of the results by the four operator groups in numerical and qualitative terms, determination of differences of opinions given by operators and members of the control group concerning the test results, comparisons of the evaluations of the documents with respect to the quality of the concept through which they were located, and finally the determination of the relevance ratio.

On the second form (Figure 10) the results were organized by individual retrieval operators for each of the retrieval tools employed and each of the questions retrieved. In Column (1) the questions were identified; a distinction is made between questions a (with) and b (without source documents); the question formulated by the operator were identified in Column (2). The number of documents recalled were entered in Column (3), the average time (minutes) spent, in Column (4). In Column (5) the recall of the source document was checked. The evaluations of the documents by the operator were entered in Block 6. Block 7 was used to enter the evaluations by Group II of the concepts; and disregarding documents under 0-concepts, Block 8 was used to evaluate the remaining documents. In Column (9) the number of pertinent documents missed during the retrieval run, but located by the member of Group II, was listed. The relevance ratio was entered in Column (10); that is, the ratio of the sum of the items rated +, = or - in Block 8 and the number of the items listed in Column (3); and the Recall Ratio was entered in Column (11), this is the proportion of the total of the +, = and - rated items in Block 8 to the total number of the relevant documents in the collection.

The subtotals and totals for the data obtained in answering the questions formulated or not formulated by the retrieval operator were computed.

This form therefore permits the determination of the following information: a. Relevance ratio; b. Recall ratio; c. the relationship between quality of results and length of retrieval time; d. the percentage of instances when the basic documents were not recovered; e. the percentage of instances when papers having greater value than the basic document were recovered; f. the relationship between quality of the concept selected and the quality of the output; g. the range of operators' and Group II's evaluations; h. the extent of bias produced by the use of the operator's own questions; i. and the differences of results caused by different retrieval tools computed for individuals as well as for groups.

The third form, the summary sheet (Figure 11), will facilitate a fast review of the results obtained by the four groups using each of the three retrieval tools. It will show the actual number of documents retrieved, the average time spent per document, the number of the basic documents retrieved as well as elimination of zero concepts, and whether the questions used were those of the operators

or not. Moreover, all pertinent documents in the collection but not retrieved during the retrieval run were listed in the "not located" lines with the required notations and ratings.

Most of the ratios resulting from information enumerated in Figure 10 will be determined more rapidly from this summary sheet.

3. THE FIRST-GENERATION ABC METHOD IN OPERATION

3.1 The Multidimensional Format

We enumerated among the various characteristics of the ABC method its capability of simulating the multidimensional format of studies that deal with different disciplines or with a variety of hierarchical levels within one or several organizational structures of science, technology or other professional endeavors. Such a claim may have appeared to be presumptuous for a number of reasons:

(a) The presentation of the ABC dictionary had linear dimensions as does any written or printed matter.

(b) Because of their linearity, most generally acknowledged schemes and methods of subject organization fail to provide adequate approaches to modern scientific and technical information, and in particular, to reports on such creative research along unforeseen border lines and within areas where disciplines that were once considered far apart now meet or overlap.

(c) Our claim was made without substantiation. Rather, we were compelled to postpone explanations and detailed proof until this time, when the dictionary of the unclassified and fairly representative test collection gives us full freedom to point out methods of approach, procedures, capabilities, and responses.

The example given in Figure 12 illustrates the retrieval operation of a scientist who approached the collection with the purpose of locating information on computer memory and switching devices. The ABC method provided him with the choice of starting his search in the dictionary under the terms: computer, switching (switches), or memory. In this instance he turned first to concepts clustered around the word "memory"; there he located the statement listed as A10 on the chart, which will be called the primary concept. Additional concepts, clustered around "memory" were easily located nearby. Each of these concepts (through the combinations of asterisk-terms and letter codes, which are not indicated in the sample) guided him directly to the card catalog where he found the complete bibliographic descriptions of the papers concerned with the well-defined, specific subjects. Because our scientist was convinced that the concept "A10": "State of the art of ferrite-core and magnetic thin-film memory device" clarified his original search problem, he used the important content terms: "ferrite-core," "magnetic" and "thin-film" as new clues when he continued screening the dictionary. During this second round, he identified 31 additional concepts (B1-11, C1-8, D1-12) as further expansions and refinements of the first very general, broad definition of his problem, and thereby increased his background information through the instantaneous access to the related, referenced literature. This retrieval process was accomplished in a very short time. Our scientist, therefore, decided to push his search into the third stage. In the "ferrite-core"

cluster (B), his attention was called to a specific detail of his design problem: "Transistor ferrite-core amplifier as logic-circuit for switching equipment (B11)." Again he followed the leads provided by this concept, searching two additional aspects: "logic-circuits" and "switching," He could have selected the latter term "switching" at the start, but omitted doing so, without lasting detriment to his task, because loops or links lead from each important cluster or facet to all others related to them. This third proliferation of the search yields 13 more concepts (E1-7, F1-6).

If we assume that our scientist abandons his retrieval efforts at this point (after he has spent about 10 minutes screening the dictionary), we can point to some of the discoveries he has made during this brief interval.

He has collected information on ferrite-core, magnetic, magnetic-core, magnetic-disc, thin-film, tunnel-diode, and superconductive memory devices, on applicable switching devices of various types, designs and characteristics. He has found leads to literature dealing with components and materials and with different principles and a variety of applications. He was guided to the subject of "logic-circuits," which was not verbally expressed in his original query, and to particular related aspects (materials, components, designs) and a bibliography on the methods of using magnetic logic circuits.

This example demonstrates the great flexibility of the system. The very specific concepts are coordinated in a broad program in a meaningful manner. In most instances, the innumerable links and loops assist in the continuing refinement of the hazy and incomplete statements of the problem and provide guides to a variety of solutions (principles, methods, and designs) and possible applications. While browsing through the dictionary, scientist and engineers adjust and supplement their original search strategies and objectives. They gain in knowledge and understanding as they screen the combinations of concepts and of facets of concepts, and perceive their problem, not only in its true scope but also in its relations to other similar parallel or overlapping efforts and disciplines (See Figure 13). Because the ABC method has the capability of tying together specific subjects and methods pertaining to different disciplines, it matches the multidimensional structure of modern science and technology, destroys the walls separating specific subject areas from each other as happens in conventional systematic arrangements, and stimulates creative work and thinking.

3.2 The Educational Capability

The second example (Figure 14a through 14e) is concerned with the search by an engineer who wanted to obtain introductory and more specific information on microminiaturization techniques for high-frequency amplifiers. The problem in his mind was general and ill-defined when he approached the ABC dictionary. He started his search under the broadest of the content terms he had included in his initial statement, the word microminiaturization.

The five concepts (Figure 14) A1 - 5 led him as could be expected to documents of such general nature and information as state of the art surveys and bibliographies (1,3,5), terminology and definitions (4), and components, devices and circuits (1,2,5). More important than the publications represented by these concepts and made available to him without delay through the ABC card catalog were the additional significant key words with which he became familiar when he rapidly scanned the concepts listed under the term: micro-miniaturization. He found in succession: semiconducting devices, molecular electronics, thin-film, and integrated circuits. He knew, of course, that these subjects were of great importance to his problem but had not included these particular terms in the original formulation of his problem. With his memory refreshed, he continued by selecting for the second round of his search the last located content word: "integrated circuit."

If we had looked over his shoulder, we could have followed him in the fast advance of his efforts as he jotted down (from Block B) concept codes leading to bibliographies (15), surveys (14), information on terms and definitions (16), general aspects of systems (10), and circuit design (18), components (12 and 17), packaging (13 and 19), and finally his main objective, amplifiers and amplifier circuits (6,7,8,9, and 11). During this scanning period his attention was called to the key word "micro-electronics" (14) and without stopping, he turned to the listings under this term in the dictionary.

This third key word (Block C) provided him with approaches to state-of-the-art surveys (25, 28, and 35); among them one on Soviet developments (29), on microelectronic systems (30), on amplifiers, his subject of primary interest (20, 22, 31, and 37); on components, active (39) as well as passive (24 , 39 and 46), on different techniques (or types of solutions) such as thin films (41 and 42), integrated circuits (35), and printed circuits (25, and 43), and on a variety of other aspects, e.g. packaging (32), interconnections (21), thermal effects (23, and 34), and reliability (26, and 33).

Because our engineer encountered the key word "thin-film" several times during his screening operations (28, 24 and 5), he decided to spend another few minutes with the concepts related to this particular subject.

The results were fast-compiled (from Block D): concepts for bibliographies (65), state-of-the-art surveys (66), circuits and circuit construction (56, 57, 59 and 63); amplifier circuits in particular (46, 50, 51, 54 and 62), passive components (44, 45, 47, 48, 53, 61, and 64) as well as active ones (52, 55, 62, 67, and 68), preparative techniques (58, 60, 64, 69, and 70) and on noise (59) and packaging (49).

Our engineer stopped his search on this point because in about 10 minutes he had gained access to and had accumulated a great number of pertinent, valuable studies and papers that inform him of the state of the art and various techniques and assist him in preparing his own better substantiated and more persuasive plan of operations.

Because it is significant for the consistency and the educational quality of the ABC method, we will outline very briefly the search strategy that would have developed if our engineer had started his search with the second broad term in his first formulation of his problem, the term amplifier.

We omit the analysis of the various types and aspects presented by the concepts concentrated around this term (in Block D), but point only to such different key-words as circuits (89), microelectronic (86) and thin film (88) to which this second approach would have guided our searcher.

Because these were the key-words he had used in his actual search process, we can conclude that an investigator with a conscious or subconscious objective in mind will be guided to identical results whatever key-word he may select as his first approach. The subject descriptors or concepts are so completely interlaced that if he picks the most general term, he will in the progress of his search encounter the more specific terminology which describes systems, circuits, devices, components, materials, manufacturing, processing and packaging methods, applications and environmental factors or any other factor closely or loosely related to the subject identified in his original formulation; and if he starts with the most specific descriptor, he will encounter the more general aspects, the principles and applications as well as the relations to other subjects or disciplines.

The method lends support not only to the memory of the retrieval operator but also to that of the person responsible for the standardization of the individual concepts at the input time. We mentioned in our first report the tools (dictionaries, thesauri, etc.) used by us to assure consistency in terminology. However, the best prepared SOP's cannot prevent the inclusion of synonyms or near synonyms unless the research analyst or the person revising the concepts remembers the established rules. While it is difficult in other systems to detect a faulty input, the ABC dictionary combines in homogeneous groups such similar expressions as molecular electronics, microminiaturized circuits, 2-D circuits, micro-electronic circuits, miniature circuits, integrated circuits, etc, thus enabling the editor not only to establish rules for standardization, but also to apply the rules (whenever they may have been occasionally forgotten for a particular input) by simple and inexpensive corrections or by introducing useful cross references into the ABC dictionary. Prior to these corrections, a mistake will not impair the retrieval because, the fabric of the system, its modes and interconnections will guide the investigator to every aspect of the analyzed and verbalized information.

4. CHARACTERISTICS OF THE SECOND-GENERATION ABC SYSTEM

Briefly, the ABC system in its present form, is based on two tools: the ABC dictionary and the ABC card file. The ABC dictionary is a list of concepts specifying the various types of information in the collection. The card catalog gives the title, accession number, etc., of each document. We must, however, emphasize that the ABC method furnishes immediate access to available pertinent literature once the proper concepts are found. It would, therefore, be logical in principle to incorporate the accession or location symbols of the analyzed documents in the body of the dictionary under the concepts to which they pertain. But it is certainly more practical to provide this information in a separate listing arranged alphabetically by codes identifying the concept. In the latter case, the investigators having selected certain concepts in the dictionary will merely note the letter code and without interruption turn to the described list of document accession numbers and use them in requesting the documents.

In the first-generation system, this listing of documents and accession numbers in the form of a card title file (organized by asterisk-term and code in alphabetical order) is not a return to another form of the conventional subject card catalog. This format was selected for two reasons: (1) to introduce the reference librarian to the ABC method, and (2) to give him as well as the subject specialist an additional opportunity of refining his request. If the evaluation of the documents was properly made, any refining could probably be made not so much on the basis of the title as on personal or corporate author, publication date, size, or abstract, if given.

In this context, the inevitable difficulties or shortcomings of the first-generation ABC dictionary and card catalog can be discussed along with their solutions, and additional refinements can be proposed. They resolve themselves specifically into the introduction of additional evaluative information to provide a further basis for selection and the redistribution of information in the ABC dictionary and the card catalog, so that the former is made a still more efficient and effective tool and the latter becomes a more meaningful part of the retrieval system.

4.1 ABC Dictionary

A coordinator and planner of activities concerned with scientific and technical information control has recently stated that "...languages, especially semantics, are not amenable to simple algorithmic representation..." and that "...any ordered pattern based upon languages or semantics must be at least as difficult to represent as the language is..."[9]

---

[9] Kelley, Jay Hilary, "The Entropy of Knowledge. Speculations Toward a Theory of Information Retrieval," 10 Nov 1964.

The truth of this observation is the vexation experienced in automating the management and dissemination of scientific and technical information. We place it at the head of this section for guidance.

If it is time-consuming and expensive to develop an adequate algorithmic representation, to organize all new information for storage by applying its signs and symbols, and to translate all requests into the same artificial language prior to any retrieval operation, if planning and programming for these input, throughput, and output activities absorb valuable manpower in great quantity, and if scientists and engineers encounter delays when their queries must be interpreted, programmed, and processed under the rules of such a system, one can logically conclude that it is desirable to search for simpler, cheaper, more direct, and more effective solutions; where financial means are lacking for the time-consuming approach, this is absolutely necessary.

It was this challenge that led to the initial development of the ABC method, to the elimination of the two-way translation problem, to the reduction of algorithms to a bare minimum, and to the utilization of natural language. A computer program was applied only to provide the tools permitting access to a collection through the medium in which scientists and engineers have learned by education and experience and have been accustomed to think, to speak and to write.

The language is the language of the handbooks and textbooks; it is the language that expresses clearly and without difficulty the true complexity of the problem at hand, depending on the number of characters that can be permitted for a given statement. Moreover, it is the up-to-date language of the specialist, adaptable to changes in meaning, to the instantaneous additions of new words and concepts, and to the elimination of the superfluous ones. Despite all this, this method facilitates standardization through automatic and complete cross indexing, and preliminary or hastily introduced terminology can be easily detected and replaced with an accepted standard form.

In addition to the improvement of standardization with respect to terminology (content words as well as function words) the automatic cross-indexing provides for the interlinking of the entire subject matter presented in the analytical concepts. Without human effort, the bridges are built leading from one discipline to another, from the specific to the general, from the most general, sometimes hazy approach to the more specific one, from theory to systems and applications, and from broad engineering aspects to the means of realization: the devices, circuits, components, and underlying principles.

This is not to say that the standardization is easy and adequate or that the system is without disadvantages. The first-generation machine program completely organizes the standardized concept-phrases thoroughly, accurately, and expeditiously in an alphabetical arrangement of all their key content words. It further provides for a secondary grouping under these keywords by alphabetizing the 60 letters after the keyword. If the preposition and the syntax

were completely standardized, we could, under a given keyword, obtain groups dealing with such homogeneous aspects as environment, application, structure and composition. However, the terminology and syntax are not so standard as to provide satisfactory grouping under keywords broad enough to include pages and pages of concepts. This standardization problem is essentially the same one encountered in the automatic machine translation of languages.

We prepared preliminary rules for the standardization of language and syntax, but recognized soon that any interpolation of adjectives and other types of qualifiers would disrupt the grouping of logically related ideas, despite the most careful planning. When the prediction is realized that "...computer adaptability will also ...include the capability of processing natural English text at a level of sophistication now possible only to humans" and when programs are available that can standardize the concepts written by the individual analysts,[10] the human efforts currently spent in this respect on the ABC (as well as on any other system's) input will simply be terminated and replaced by automation.

Because extensive studies are being undertaken and programs are being prepared for the standardization of English texts to facilitate machine translation into foreign language, we assumed that the results of these endeavors would assist us in bringing the concepts prepared by a number of experts into a consistent and useful format through alphabetic ordering under keywords.

We discussed our problems with experts working on the automatic generation of standardized texts, but concluded that we could not and should not burden our already difficult task with complex and extremely costly programs which, according to the best estimates, would not become operational before the end of a decade. We could not afford the luxury of investing large sums in fascinating, but still unpredictable investigations; and, still more important, we could not postpone our solutions for many additional years. We decided that what can be accomplished should be accomplished right away; and this in as simple, practical, and economical manner as possible.

We recognized that certain disruptions introduced by automatic alphabetization had to be eliminated, and concepts that rapidly accumulated in the ABC dictionary under broad and significant content words such as amplifiers, antennae, diodes, lasers, oscillators, plasma, transistors, etc. be organized for rapid and easy location.

Therefore, work was undertaken to prepare a practical, flexible scheme for grouping in subdivisions the information under the

---

[10] Simmons, Robert F; Sheldon Klein; Keren McConlogue. Co-occurrence and dependency logic for answering English questions. System Development Corp., Santa Monica, Cal., Apr 3, 1963. SP 1155, p. 1.

different important keywords, to prepare a program capable of listing the same concept under at least three of the logical subdivisions of such a special superposed scheme whenever desirable and to automate all clerical functions, such as the reproduction of the required numbers of concepts, the filing of the concepts into the various subdivisions, and the printing of the organized sections and subsections of the ABC dictionary.

Under the present plan, the number of subdivisions in one given scheme is limited to 676 because of the two-letter code used for identifying them in the machine program. These codes (alphabetically arranged) for each subsection into which it is to be inserted are added to the respective keywords of a concept.

The machine program will insure the following computer operations: 1) alphabetization by keywords; 2) recognition of the different codes attached to them; 3) reproduction of the required number of concepts; 4) arrangement of the concepts by the code symbols; 5) insertion of the headings ahead of the subdivisions (corresponding to the code) from a second tape; and 6) printout of the subheadings and concepts in order and eliminating the codes from the printout.

With the assistance of Dr. Louis dePian subgroups or subject schemes (microschedules) for 30 different keywords have been introduced.

To illustrate the improvements we can expect from this change, we analyze a small number of the concepts organized by the content word "amplifier" as they were published in the first-generation ABC dictionary (Chart II). The deficiencies are quite apparent.

The alphabetized keywords followed by a comma, period, equal sign, asterisk, etc. are separated from those without subsequent marks or symbols. Whether the function word "to" introduces an infinitive or serves as a preposition is not considered; it is therefore a link between quite unrelated concepts. Because the machine program is geared to effect alphabetization behind the keyword (that is, to the right of the break), the reader's eye will get used to scanning this segment of the line exclusively in trying to find appropriate concepts, and will tend to miss the concepts that end with the keyword. With the introduction of microschedules, this disarray of individual concepts caused by mechanical alphabetization and inconsistencies in standardizing concept terminology and syntax is eliminated and replaced by the subject organization in Figure 15. For example, the concepts organized under the key term, Amplifier, in Figure 14 will be presented in the following format:

Amplifiers - general

Multistage -

Concept 93

Amplifiers frequency
Radio frequency

Amplifiers-function

Intermediate frequency -

Concepts 94, 99

Wideband -

Concepts 91, 96

Linear -

Concept 95

Low-noise -

Concept 87

Band-pass -

Concepts 89, 90, 97

Amplifiers-mode of operation

Tuned -

Concept 90

Parametric -

Concept 98

Amplifiers- active element

Tunnel - diode -

Concepts 87, 88, 92

The individual concepts would be printed out fully under the subdivisions alphabetically as they presently are.

It is obvious from this example that the organization of the second-generation dictionary will greatly simplify and speed up all retrieval operations. The additional encoding will increase the input cost slightly. However, there is no need for also encoding the inquiries, a requirement for retrieval from a collection organized by a coordinate indexing system and burdened by the additions of roll and link indicators.

Two additional format changes are designed to improve the appearance and usefulness of the dictionary. First of all, we will eliminate the length restriction of the individual concept to one line. A new program in preparation will accommodate concepts of any length. Second, we will print the concepts in a different arrangement (Chart III as an example; a final choice has not yet been made), with double printing to produce a bolder type face for the headings.

Although the secondary organization of the concepts around the key term will be mainly accomplished by superposed subject schemes, we will continue working on the refinement of our rules for concepting. Something other than the rule that an overall concept must be prepared to tie together the various different concepts assigned to one paper is needed. In addition, we will give considerable attention to the standardization not only of the terminology, but also of the syntax as soon as practical results are available through the research on generative grammar and automatic translation methods.

4.2 Second-Generation Card Catalogs

Another major objective of the second-generation ABC method is the reduction of the dictionary to the smallest possible size by moving some information to the card catalog. The major items scheduled for this operation are parameters and descriptive information; e.g., form of the publication, its level of audits difficulty and its method of approach and phase of research.

4.2.1 Parameters

In discussions of engineering data, the main emphasis is generally placed on the selection of manufactured materials, components and devices; that is, shelf items which meet stated requirements for

the operation of particular systems or subsystems under development or in production. Provisions are rarely made for scientists and research engineers in search of information on the performance capability of available hardware, let alone items still under study or in the development phase. HDL research personnel have frequently insisted that the contents of scientific and technical reports and published letters be analyzed and organized for access by such parameters as frequency, power, current, voltage, particular environmental factors, efficiency, etc. to facilitate the construction of prototypes and models incorporating the best or latest components.

We could not ignore these eloquent requests because we recognized that in the near future any efficient retrieval systems for scientific information must also possess the capability of fitting into an overall scientific environment and of supplementing the reference systems for engineering data by supplying scientists and engineers with the data of not yet completed items.

By way of form, Mr. M.M. Algor, an electronic engineer at HDL, suggested a simplified notation of parameters. The numerical values, for example of a frequency, usually expressed as $F = n \times 10^k$ Hertz (n being a number composed of two or more significant digits, k the appropriate, either positive or negative, power of the base 10), would be transformed to F = n (k). Finally this would be abbreviated to FnPk or FnNk where P and N denote a positive or negative value of the exponent and, in addition, separate the digits of n from the numerical value of the exponent k . He also pointed out that the same method would make it possible to encode any numerical parameters at the cost of no more than 7 digits or characters. His suggestion was applied to the writing of concepts (Chart IV).

Such significant parameters as frequency (F), current in amperes (A), magnetic field in gauss (b), energy in electron-volts (E), power in watts (P), acceleration in g's (G) etc. were identified and their code symbols (combinations of capital letter and numerals) were inserted with concepts either like adjectives in front of the term they qualified or at the end of the entire concept to which they pertained.

Although various methods of simplification and standardization were suggested during conferences between HDL and GWU, many desirable changes had to be postponed to avoid further delay in the publication of the ABC dictionary used during the test.

At this time a detailed SOP has been drafted to govern the inclusion of parameters and their numerical values into our information retrieval program. The parameter designations and the basic units to be used have been determined, and the presentation of numerical values has been standardized.

However, the inclusion of the parameters in the ABC dictionary as well as the form entries (e.g. bibliography, survey, etc.) is undesirable as may be seen from the following: Assume we continued to combine the parameters with the concepts and provide access to them through the ABC dictionary. If four different parameters are added to the average concept composed of five permutable key words, these four additions will create four new concepts or the requirements for the alphabetizat' of six key terms four times; therefore, 29 lines will be print.d our instead of five and the size of the dictionary will be increased by a factor of six. Because of the current rate of growth and the need for supplements to and accumulated editions of the ABC dictionary, this method would result in the production of very bulky and also repetitive reference tools, as not only the added parameters but also the (five original) contents terms will be rotated and reproduced each time.

The increase in computer time and cost of reproduction will be in proportion to the increased size of the dictionary, but the difficulty of organizing and using the dictionary will also be increased considerably. We have, therefore, decided to exclude the parameters from the dictionary and to print the information on catalog cards. The parameter notation will be found at the bottom of cards which constitute the so-called ABC card catalog, that is, the file where the searcher locates the titles under the concepts he has selected in the dictionary. Furthermore, the same information will be printed out on the top of additional title cards and made accessible in a separate file through the following arrangement: Cards regarding a specific parameter will be subdivided in alphabetic order by names of systems, devices, instruments, components, etc., and these subdivisions organized by the numerical (parameter) data filed in ascending order. For example, behind the guide card for the parameter, frequency, we will find such subject headings as: Amplifiers, Antennae, Diodes, Oscillators, Switches; and under each of these subject headings the pertinent titles on separate cards filed in numerical order according to the frequency (ranges of frequencies will be represented by two cards, one to be filed by the minimum and the second by its maximum value).

4.2.2 Descriptive Information

By the formulation of very specific and complex concepts, the ABC method facilitates direct access to correspondingly specific and complex information. This versatility was also used in the first-generation dictionary to advise the user of the types or forms of analyzed publications in the collection; if they were analytical or title bibliographies, collections of papers (symposia or proceedings); if they covered a small period, a major segment, or the complete entity of a particular research or development project; if they reported on a theoretical study, an experimental investigation,

or a test; and if they were addressed to the student, the generalist, or the specialist.

These very detailed descriptions by form and type of publication, and by level of treatment led to the specific entries in the ABC card file. Because the use of the card file was not eliminated and the development of a more condensed and efficient dictionary was desired, we decided to provide the information concerning the format in a different and economical manner. In the future, the second-generation dictionary will convey only subject matter in most precise statements or concepts; but on the individual title cards of the ABC card file, letters will be added to the shelf number in order to denote whether the particular title represents a bibliography (b); a collection (symposium or proceeding) (c); a theoretical (t) or experimental (e) study; or a progress (p) or summary (s) report. These symbols will be determined at the time the concept is being prepared or standardized; and they will be used to establish and print out subheadings under the respective asterisk terms in the card catalog and to guide the retrieval operator in his selection. These are not to be confused with the subheadings in the dictionary, some of which are being carried over (into the microschedules), to be used to aid in searching for the appropriate concept.

5. CONCLUSION

5.1 Second-Generation ABC Method

This report is an introduction to the presentation of the test data and evaluations, which will be published in a few weeks. In these concluding paragraphs, we wish to emphasize the characteristics of the second-generation ABC storage and retrieval method, sum up the basic problems we encountered in planning and performing the test, and call attention to some areas requiring further investigation.

The second-generation ABC system is a hybrid in as much as subject classification has been superposed on the otherwise predominantly alphabetical arrangement of ABC concepts. Both types of organization supplement each other most effectively. The alphabetic arrangement of the standardized concepts has the great advantage of preparing a large amount of information for fast and meaningful access, because it is "the most highly successful ordering pattern based upon tradition with very little components of natural order or logical order."[11] Its disadvantage (at least with the available program) is its inability to furnish logical order in the specialized subject areas, that is, under the key words. This deficiency is overcome by the subject classification in these areas, which should be distinguished from systematic schemes covering large subject areas. Because they are pre-conceived and pre-prepared, all comprehensive classification schemes are difficult to adjust; and because they are linear in structure, they lack the flexibility of incorporating the new aspects and the new disciplines of the rapidly expanding science and technology. They separate as much subject matter as they bring together. However, in a very narrow and specific field (under a keyword) they permit a logical organization that is acceptable and helpful.

Through the ABC dictionary and retrieval method, an interface has been established between the scientist or engineer and the contents of the collection. The investigator is no longer required to discuss his problem with a documentalist and to assume that his problem is correctly and fully understood; or to possess "the a priori knowledge... of that which he is requesting in order to produce a successful output."[12] When he opens the dictionary with a general or incomplete formulation of his requirement, the contents of the collection are not only displayed before him in a language he fully understands, but the multidimensional characteristics of the presentation lead him to more precise definitions, to a clearer recognition of his task and its complexity, and to information as broad or as specific and as theoretical or as practical as he may desire. The system itself serves an educational purpose, in that it gives the researcher more than he brings to the information

---

[11] Kelley, Jay Hilary. "The Entropy of Knowledge. Speculations Toward a Theory of Information Retrieval," 10 Nov. 1964, p. 39.

[12] Ibid.

office, and at the same time, enables him to develop and adjust his own search strategy and to select those areas or subjects he wishes to cover whether close to or removed from the original formulation of his question.

The retrieval operation using the second-generation ABC dictionary is graphically shown on Figure 15A.

Since the investigator or searcher makes his own selection from the collection, he obtains the information he needs faster and at a lower cost. In Figure 16, the ABC retrieval process is compared with a characteristic retrieval operation performed on the basis of a coordinate indexing system. In this latter process, the questioner with his problem in mind must first face a documentalist, who will prepare the search strategy and the program as soon as he believes that he has understood the investigator's question. The program is then punched, inserted into a computer to recall the corresponding information from its memory, and in most cases, the print-out is checked for adequacy and pertinency by a subject specialist prior to its release.

The ABC retrieval method on the other hand necessitates only the following steps:

1. The screening of the dictionary for the identification of the appropriate concepts (by the asterisk-term and letter code);
2. The inspection of the card catalog under the term-code combination and the withdrawal of the pertinent documents by accession number from the shelf.

It should be emphasized in this connection that an increase in the effectiveness of the system through screening the dictionary more intensively will not affect the efficiency of the operation to a proportional extent.

The directness and simplicity of the manual retrieval operation, the organization by parameter, type, and level of difficulty in two ABC card files, and the automatic print-out of bibliographies by the concept number as well as by the subject-groups are other gratifying improvements of the second-generation ABC method.

Moreover, it should be considered an advantage that the system discourages the ignorant from participation in the input procedures.

While in other systems the indexer may pick conspicuous terms from the title page, the content table, or the body of the paper and insert them (after a process of standardization), appropriate concept phrases can be prepared only by those who are capable of analyzing and judging the substance and quality of the papers. The results of a selective input and meaningful approaches will enhance the quality of the output and decrease the cost of the service.

And last but not least, to those agencies not encumbered by traditions and by costly investments of repetitive studies and

surveys, complex and time-consuming programs and procedures and inappropriate machinery, the method offers relatively simple, flexible, and apparently economic solutions to the problem of displaying current scientific and technical information in an understandable, meaningful manner for selection by the scientist and by the engineer at the work bench as well as in a supervisory, planning, or executive position.

5.2 Characteristic Aspects of Test Preparation and Performance

The design of the test and its procedures for the performance and evaluation proceeded in the presence of various major difficulties. To conduct the test to obtain meaningful data, we had to create a true-to-life situation. In addition, since the objective of the project was to evaluate the system as such and not its operators, we had to eliminate the human error factor at least to the extent that it might exceed the permissible or expected tolerance; and we had to weight the distortions caused by human subjectivity at every step of the test activities: the selection and compilation of the collection, the preparation and standardization of the concepts (or subject approaches), the formulation of realistic and pertinent questions, the test performance consisting of the identification of the significant concepts and the appropriate documents, and finally the evaluation itself.

A further complication was added in that this test was not separately funded, but (with the exception of the information analysts) had to be conducted by "volunteers" who under the pressing burden of their main assignments in the laboratories did not always cherish their role. For many of them, the test provided the first opportunity for an actual contact with the ABC retrieval method.

Because these difficult problems and situations were recognized, we introduced as many controls as possible to take advantage of the disadvantages of the situation.

Size, scope, complexity, quality and currency were the determining factors for building a test collection to give meaningful and representative responses to realistic questions.

The controls used for the formulation of realistic questions were twofold: 1) some questions were generated on the basis of papers randomly selected from the collection and others on the basis of only a general knowledge of the scope of the collection; and 2) the questions suggested by one group of scientists and engineers were screened, evaluated, reduced in number, and standardized by a second group of research supervisors and administrators. We do not claim that this procedure was entirely successful;

but we preferred it over the suggestion that questions actually submitted to the library by laboratory personnel during the last year should be used because a check indicated that these verbalized questions seldom reflected the true requirements of the investigators.

The retrieval operations themselves were performed with a variety of control measures, by three different groups: 1) the scientists and engineers who had formulated the questions; 2) the research analysts or producers of the concepts; and 3) the HDL librarians (reference librarians and catalogers). We divided the first group (scientists and engineers) into two sections as a control factor: to determine what bias resulted from retrieving one's own question.

To determine the difference in details required by experts and generalists, we tested two different ABC dictionaries. A KWIC title list was used as an additional control factor. In this way, we obtained 12 test data for each question. The sequence of tools used was altered to determine the possible bias produced by the sequence. To provide for an objective assessment of retrieval, the quality of the paper on which the test question was based was used as the standard of measure; for the freely-styled question the quality of the retrieval was determined by its relevance to the query.

The adequacy or inadequacy of the concepts selected by the various operators from the ABC dictionaries was evaluated in the following manner: 1) the various combinations of individual elements contained in a given question were rated on how well they reflected the intent of the problem: excellently (+), adequately (=), still usefully (-), or inadequately (o); 2) these a Priori established combinations and their ratings were then applied to measure the usefulness of the selected concepts. While subjectivity may obscure the lines that separate the different ratings, the sliding scale makes it possible to identify the concepts that are completely worthless and unrelated with regard to the question. The number of unusable retrievals will be used to determine the extent of operator error and will provide for an error-weighted relevance ratio.

If the test data should offer a sufficiently broad base, a study will be initiated to determine the qualifications or characteristics of a good retrieval operator. It would be based upon two samples of 10 operators each: one for those who had been consistently successful, the second for those who had consistently failed. By selecting certain factors such as education, length and type of experience, papers and reports published, patent disclosures, etc., and by preparing one rating scale for each of these factors, we would produce profiles. With the sum of the weighted factors, grades for each individual and frequency curves may facilitate an answer to our question.

APPENDIX A

The Cost Factor

To comply with a number of requests, one of which was included in a review of the first report, we have kept an account of the money spent to organize the test collection in accordance with the ABC storage and retrieval method.

We are aware that absolute cost data are of no significance unless they: 1) are related to the quality and utility of the service; 2) can be translated into price scales that prevail in different countries and localities; and 3) can be adjusted to fit particular requirements. In every instance one must also consider the cost of other methods, especially the current method one wishes to replace, since a comparison can be meaningful only if an eventual increase in cost can be measured in terms of resulting improvements.

We have, therefore, reduced all expenditures to unit cost, that it cost per title; and wherever feasible, given an indication of the time involved in the individual operation.

a. For the selection of the test collection and the preparation and standardization of the concepts for 3650 accepted titles - total cost was $10,674.60 and the unit cost	$2.91
b. For the input into the computer memory we required an average of 6 punched cards per title. At a unit cost of $0.07, the cost per title was	0.42
c. For the print-out of 3 different catalog cards and one bibliographic listing of the collection, a total of 24 lines per title at $1.00/minute (1410) machine rental, the cost per title was	0.11
d. For the KWIC title list (a non-essential tool for the customary reference service) about 5 lines were permuted at $8.00/minute by way of the 1410 computer with a per title cost of	0.06
e. For permuting 4000 concepts with 7094 computer, the total cost was $150.00, and the cost per title	0.04
f. For the printing of the ABC Dictionary with the 1410 computer, the cost for an average of 6 lines per title amounted to	0.01
TOTAL	$3.55

NOTE: Additional cards create a cost increase of up to $0.02.

In the expenditures we included only the cost of printing one dictionary. Every additional accumulation requires another printing of the same title, at a title cost of $0.01. If therefore the average title added to the collection will be published in a second accumulation during the first year, and then be included in the yearly accumulations in its second and third year, the cost of three accumulations must be added, an increase of $0.03 per title.

APPENDIX B

Divisions of the Classified Catalog

A. Applications of Solid State Devices
   - A0. General
   - A1. Communications
   - A2. Computers
   - A3. Power Applications
   - A4. Control Applications
   - A5. Instrumentation
   - A9. Other Applications

B. Basic Solid State Device Circuits
   - B0. General
   - B1. Amplifier
   - B2. Oscillators
   - B3. Switching Circuits
   - B4. Signal Convertors
   - B5. Wave Generators
   - B6. Pulse Circuits
   - B9. Other Circuits

D. Solid State Devices

K. Semiconductor Device Measurements
   - Ka. Diode Measurements
   - Kb. Transistor Measurements

R. Conductive Devices

a. Diodes and Rectifiers
   - a0. General
   - a1. Point Contact Diodes
   - a2. Junction Diodes
   - a3. Area Contact (Metallic Rectifiers)
   - a4. Surface Barrier Diodes

b. Transistors
   - b0. General
   - b1. Point Contact Transistors
   - b2. Junction Transistors
   - b4. Surface Barrier Transistors
   - b5. Field Effect Transistors
   - b9. Grain Boundary Transistors

c. Functional Units

d. Magnetoelectric Devices
   - d1. Hall Effect Devices
   - d2. Magnetoresistive Devices

e. Other Conductive Devices
   - e1. Resistors
   - e2. Symmetrical Varistors
   - e3. Cryogenic Devices
   - e5. Negative Mass Devices

f. Photoelectronic Devices
   - f0. General
   - f1. Photoconductive Devices
   - f2. Photodiodes and Phototransistors
   - f3. Photovoltaic Devices

g. Luminescent Devices
h. Other Photodevices
  h1. Photogenerators
  h2. Quantum Convertors
  h3. Optical Filters
  h4. Polarizers
T. Thermal Devices
  k. Thermistors
  m. Thermoelectric Devices
  n. Other Theroelectric Devices
H. Magnetic Devices
  p. Ferro and Ferrimagnetic Devices
  p0. General
  p1. Altenuators
  p2. Isolators
  p3. Phase Shifters
  p4. Circulators
  p5. Amplifiers
  p6. Logic (Memory) Elements
  q. Paramagnetic Devices
  q1. Masers
  r. Other Magnetic Devices
E. Dielectric Devices
  s. Ferroelectric Devices
  s0. General
  s1. Electromechanical Transducers
  s2. Memory Cells (Storage Elements)
  s3. Amplifiers
  t. Other Dielectric Devices
  t1. Fixed Capacitors
  t2. Variable Capacitors
  t3. Space Charge Limited Dielectric Devices
G. Other Solid State Devices
  u. Superconductive Devices
  u1. Cryotrons
  u2. Crowe Cells
  v. Electromechanical Devices
  v1. Piezoresistive Devices
  w. Magnetomechanical Devices
  w1. Magnetostrictive Devices
  z. Miscellaneous Devices

APPENDIX C

June 10, 1964

TO: Distribution

FROM: B.M. Horton, Technical Director

RE: Cooperation in Test of Indexing System

1. The Defense Department has asked us to run a test of the effectiveness of our indexing system for information retrieval. The test is under the general supervision of Dr. B. Altmann, our Technical Information Officer, but he will need your help.

2. I ask each member of the Editorial Committee to cooperate as requested in the selection of test questions and in the evaluation of test results.

3. I would like for each Laboratory and Division Chief to cooperate by assigning subject matter experts as requested to help the Technical Information Office and the Editorial Committee.

4. With adequate participation, each person involved will need to devote to the task not more than six hours during a two-week period.

5. Area Intelligence Information Officers will arrange details of Laboratory/Division cooperation.

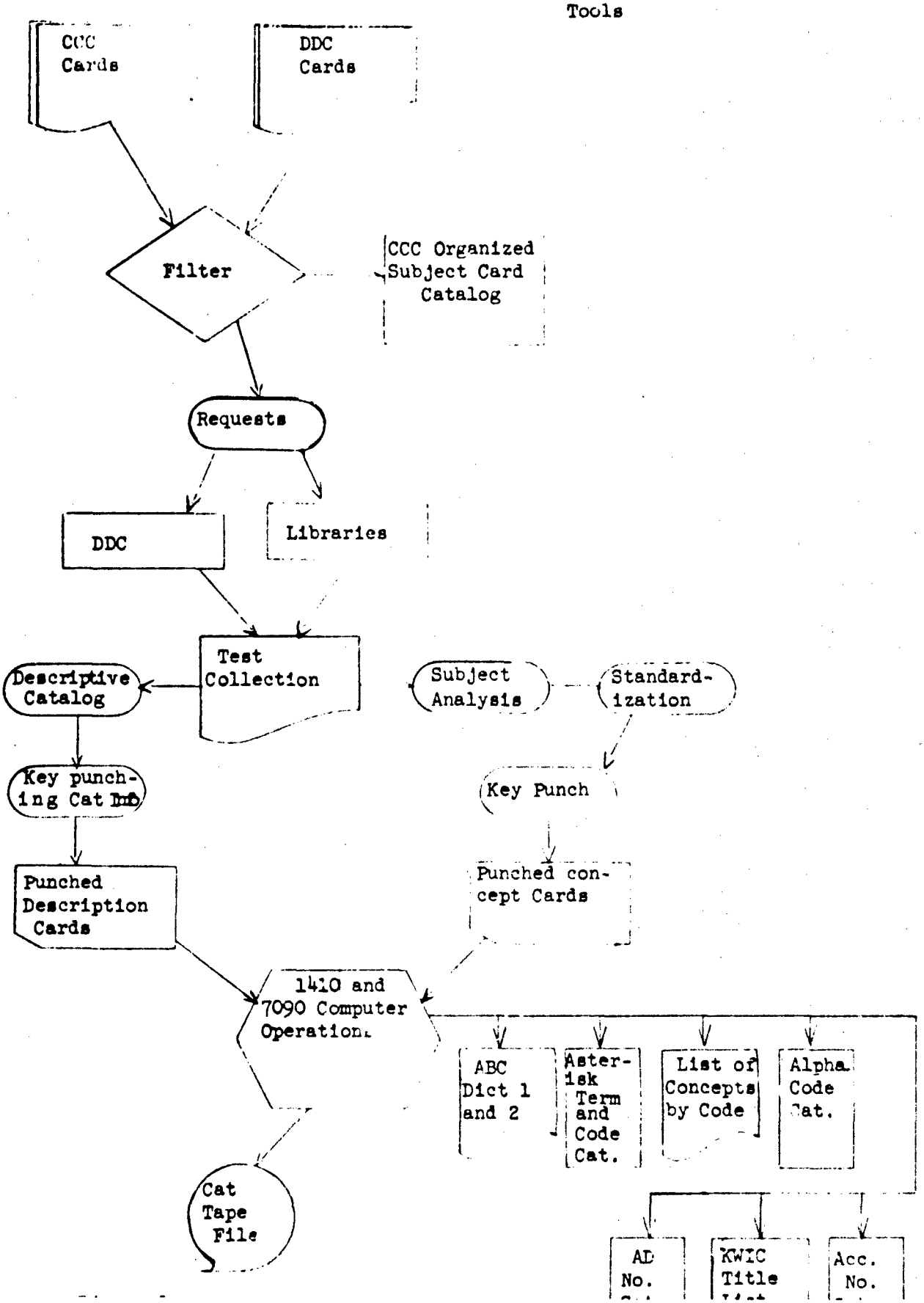
6. The test will be run on a sample of 400 documents randomly selected from a collection of 4,000 documents in the field of solid state devices, circuits, and their application. HDL personnel are now preparing questions based on these documents. These questions will be put to the index of these documents prepared by the HDL system; the documents retrieved as a result of this process will be evaluated for relevance and coverage as a measure of the effectiveness of the indexing and retrieval processes.

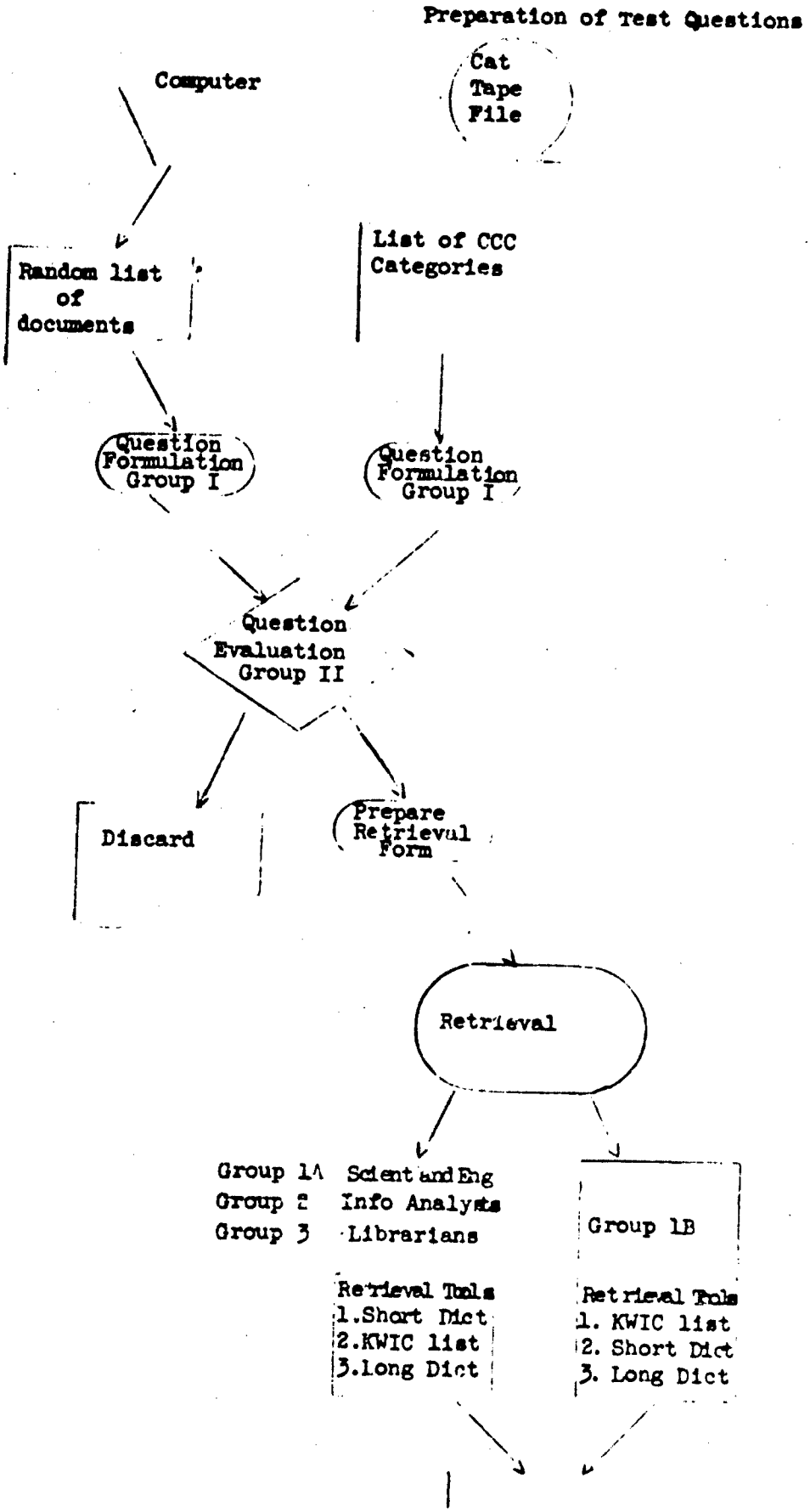
B.M. Horton

IR/bwh

Distribution: Lab and Div Chiefs: Hardin; Sommer; Hatcher; Hoff; Nilson; Flyer; Campagna; DeMasi; Landis. Editorial Committee: Eichberg; Godfrey; Moorhead; Distad; Vorkink; Bryant; McCoskey; Kalmus; Dr. B. Altmann. I. Rotkin.

Preparation of Test Collection and Retrieval Tools




Preparation of Test Questions
Computer
Cat Tape File
Random list of documents
List of CCC Categories
Question Formulation Group I
Question Formulation Group I
Question Evaluation Group II
Discard
Prepare Retrieval Form
Retrieval
Group 1A Scient and Eng
Group 2 Info Analysts
Group 3 Librarians
Retrieval Tools
1.Short Dict
2.KWIC list
3.Long Dict
Group 1B
Retrieval Tools
1. KWIC list
2. Short Dict
3. Long Dict

GUIDELINE FOR QUESTION FORMULATION

1. There are three sheets of 20 titles each attached. These titles have been selected from a listing of 400 such titles chosen randomly from the entire test collection of 4000. They are arranged so that beginning on the first page, the titles cover the area that you indicated an interest in. Since the requirements of the program make it necessary that each individual prepare at least six questions, please select four titles from the first page and the remaining ones from each of the following pages.

2. The questions should be written in such a manner that:
   a. They are answerable by the article.
   b. They are limited to the major findings, methods, devices, parameters, etc. discussed in the article.
   c. They do not cover general introductory or historical material.
   d. They are not based on footnotes or material incidental to the major topic of the article.

3. The articles used in this portion of the test are set aside in a special section of the library and can be located by consulting the librarian.

4. The last three weeks in June will be allocated for the formulation of these questions.

5. When you have prepared the questions, please submit them to Miss K. Rydlewicz, Room 207, Building 92.

6. The following is a sample of the form to be used in preparing the questions. These forms will be placed on the circulation desk at the entrance to the library.

TIO Records Information (leave blank)

Question:

Document Identification: (periodical name, volume, date, and page)

Name: Branch: Ext:

Evaluation of Question: (leave blank)

Figure 5

Evaluation of Questions

Question:

Document:

Criteria for evaluation

1) Question should be concerned with significant aspects of the paper.
2) Several questions closely related in substance or differing only in terminology should be combined.
3) Question requiring modification should be rephrased.

Approved Formulation:

Assigned to: Branch: Ext:

Accession numbers of other documents which should be retrieved. In comparison with basic document:

Of Greater Value	Equal in Value	Of Marginal Value
1.		
2.		
3.		

Figure 4

Retrieval Form

Name: Br: Ext:

Question:

Basic Document:

Time of Start: Stop:

* Term	Code	Accession Nos.
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

Figure 5

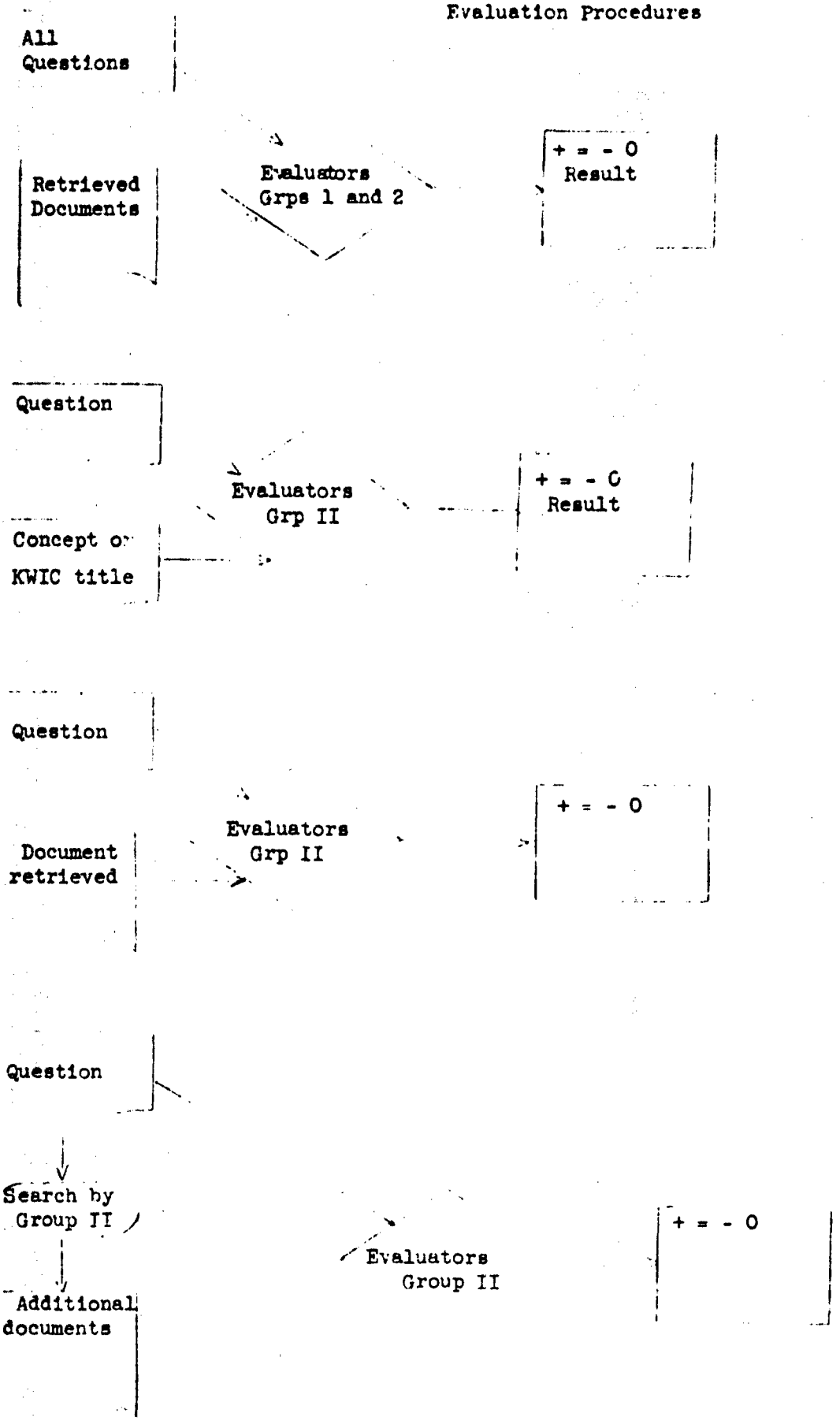

Evaluation Procedures
All Questions
Retrieved Documents
Evaluators Grps 1 and 2
+ = - 0 Result
Question
Concept or KWIC title
Evaluators Grp II
+ = - 0 Result
Question
Document retrieved
Evaluators Grp II
+ = - 0
Question
Search by Group II
Additional documents
Evaluators Group II
+ = - 0

Figure 6

Evaluation of Test Results

Name:

Group: Team:

Question:

Basic Document:

Titles Retrieved (For Group I and II)					Pertinent Titles Not Retrieved (For Group II)				
Accession Nos.	=	+	-	0	Accession Nos.	=	+	-	0
1.					1.				
2.					2.				
3.					3.				
4.					4.				
5.					5.				
6.					6.				
7.					7.				
8.					8.				
9.					9.				
10.					10.				

Figure 7

RETRIEVAL LOOPS

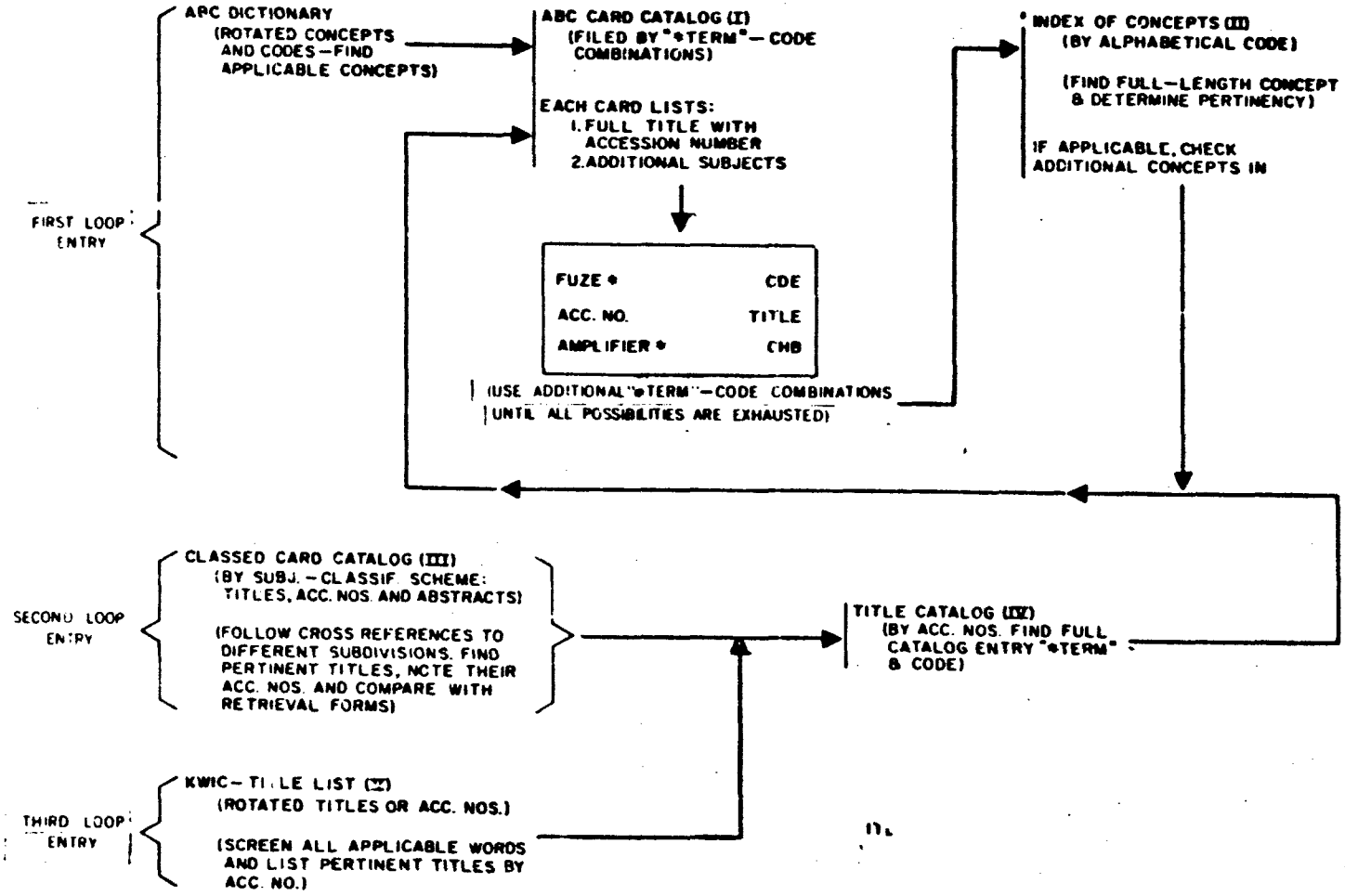


FIGURE 8

STATISTICAL ANALYSIS (Form Sheet for Questions)

Question: _____

Basic Document: _____ Basic Concept: _____

Group	1A				1B				2				3			
Evaluation by:	1A		Group II		1B		Group I		2		Group II		3		Group II	
	N	+	=	-	0	+	=	-	0	N	+	=	-	0	+	=	-	0	N	+	=	-	0	+	=	-	0
Short Dict.																											
1) own ques.																											
2) other's																											
3) concept																											
Long Dict.																											
1) own ques.																											
2) other's																											
3) concepts																											
KWIC List																											
1) own ques.																											
2) other's																											
3) concepts																											

Figure 9

Statistical Analysis (Form for Retrieval Operators)

Name:

Group: Background:

Tool:

2. own question?	3. Number of retrieved documents	4. Average time (min.)	5. Source document	Evaluation of Retrieval Runs												9. Additional documents located by Group II	10. Relevance	11. Recall
				Evaluation of document by operator				Evaluation of concept by Group II				Final Evaluation of document by Group II						
				+	=	-	0	+	=	-	0	+	=	-	0			

Statistical Analysis (Form for Overall Summary)

Groups	IA						IB						II						III					
	n	t	=	+	-	0	n	t	=	+	-	0	n	t	=	+	-	0	n	t	=	+	-	0
Dictionary I																								
Loc with 0 con																								
Loc w/o 0 con																								
not located																								
own group																								
Loc with 0 con																								
Loc w/o 0 con																								
Loc with 0 con																								
Loc w/o 0 con																								
Dictionary II																								
Loc with 0 con																								
Loc w/o 0 con																								
not located																								
own group																								
Loc with 0 con																								
Loc w/o 0 con																								
other group																								
Loc with 0 con																								
Loc w/o 0 con																								
KWIC List																								
Located																								
Not located																								
own group																								
Located																								
Not located																								
other group																								
Located																								
Not located																								

Figure 11

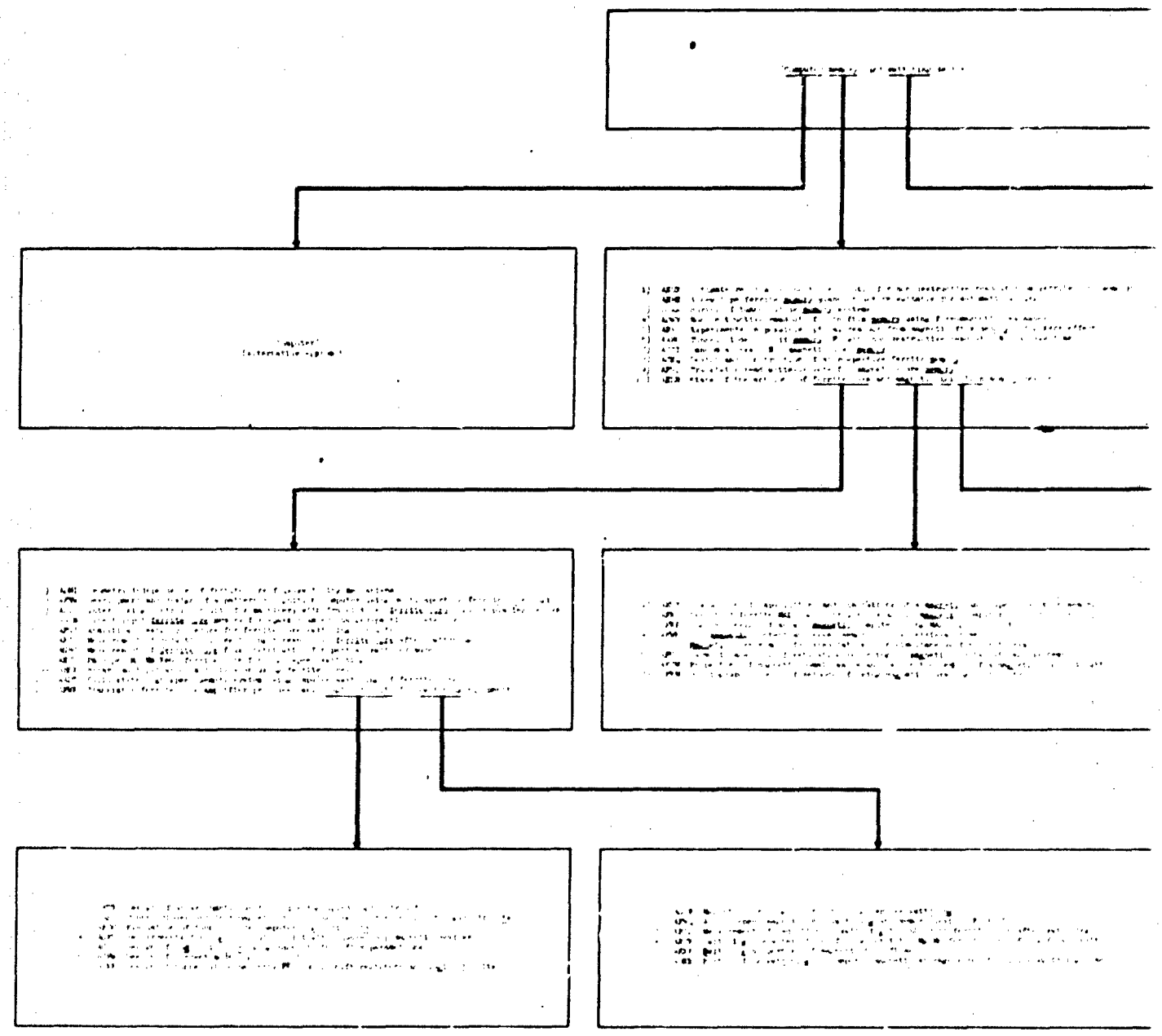
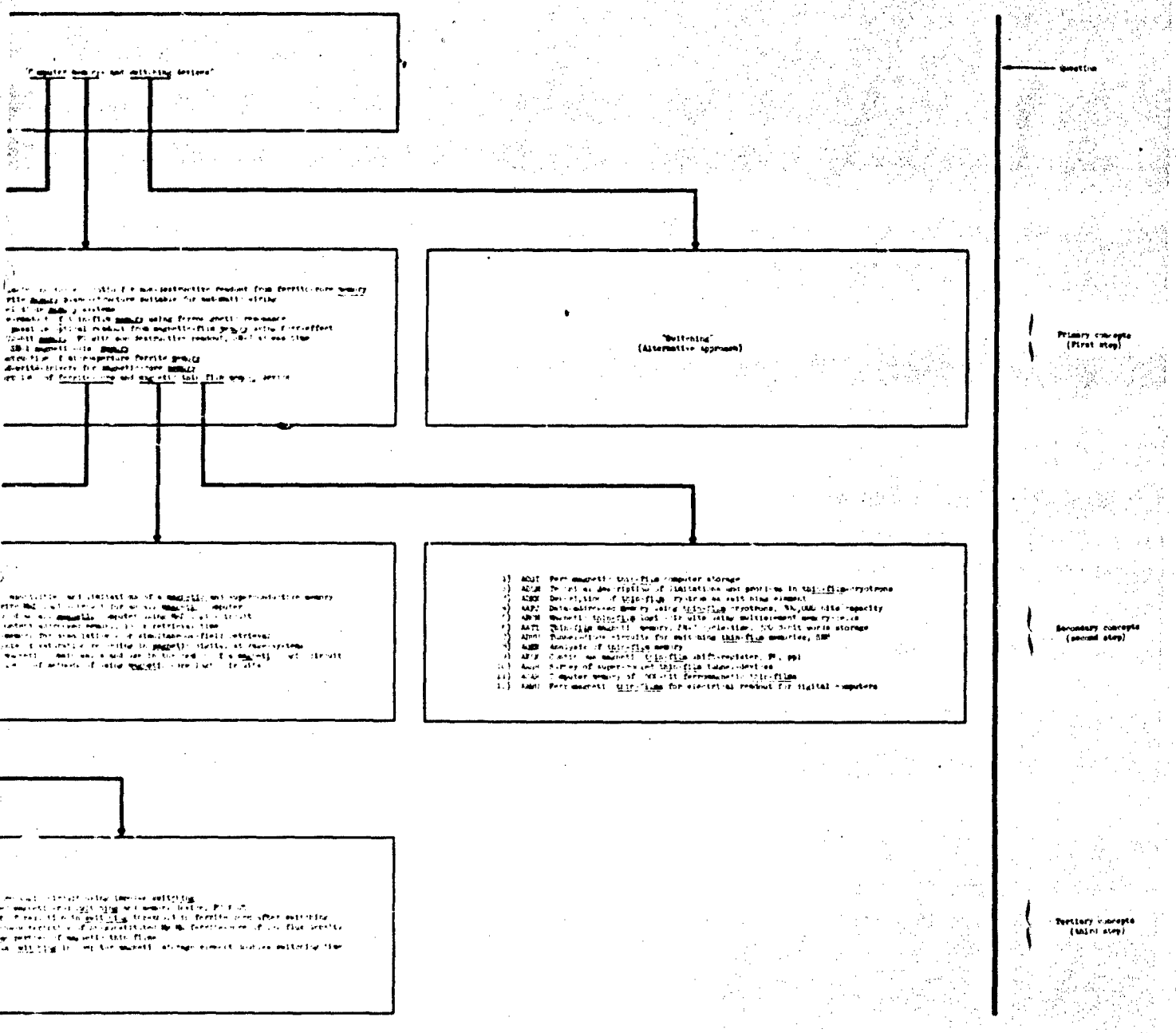


Figure 12

A


"Switching"
(Alternative approach)
Primary concepts
(First step)
Secondary concepts
(second step)
Tertiary concepts
(third step)

B

FIRST-LEVEL PROLIFERATION OF INQUIRY BY INTERRELATIONSHIP OF CONCEPTS

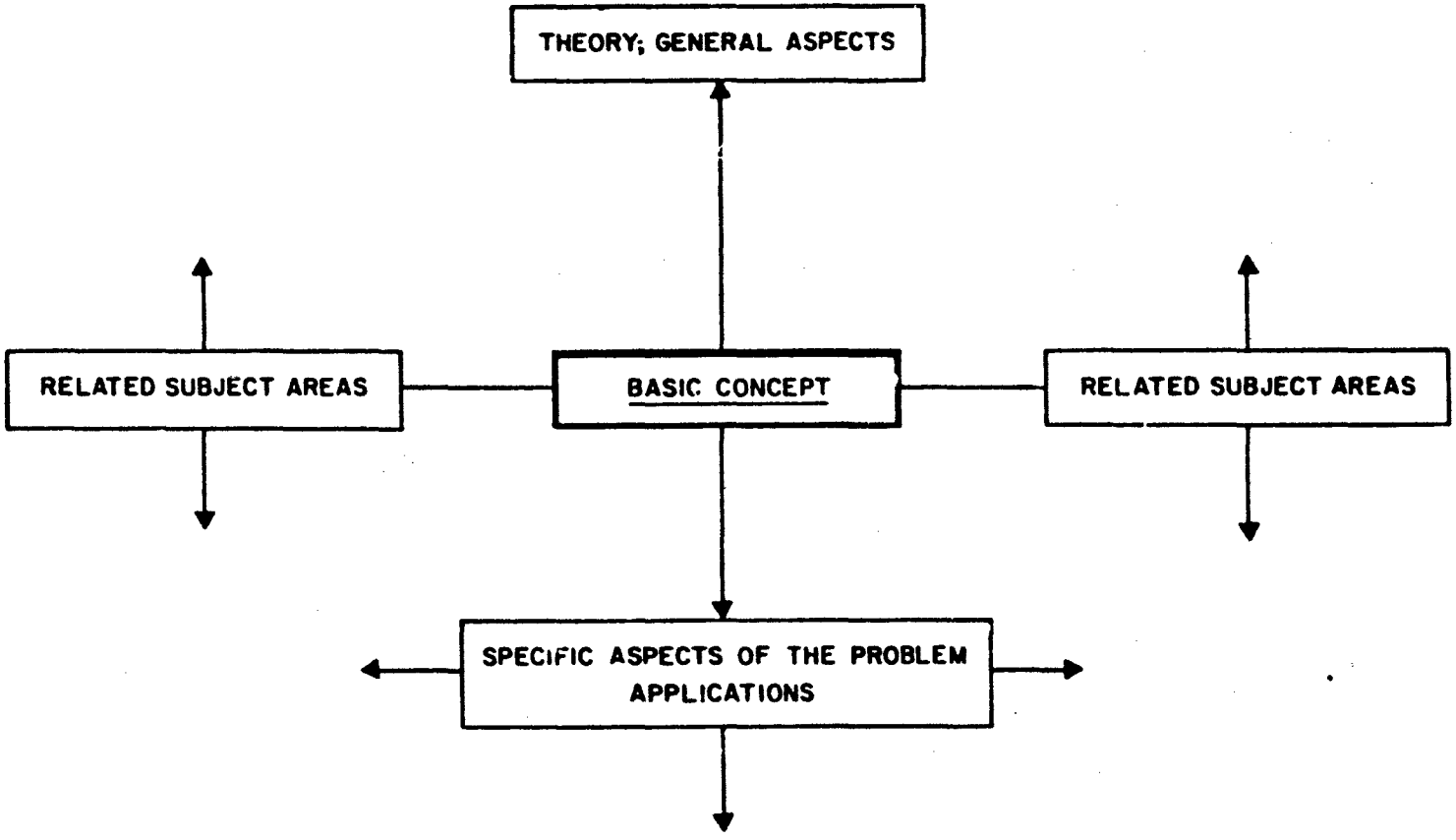


Figure 13

Question: Microminiaturization techniques for HF - Amplifiers

1. Approach: concepts clustered around "microminiaturization"

Alternative 1. Approach: concepts clustered around "amplifiers"

see Figure 14e

A

1) AECH Survey 1963 of microminiaturization-components and design
2) ADVH Limits on microminiaturization of non-redundant semiconducting devices
3) AABI Status, definitions, bibliography 1962 of microminiaturization, molecular electronics, etc.
4) ADMG Terms and definitions used in microminiaturization and integrated circuits
5) ADSM Annotated bibliography 1963 of microminiaturization, thin-films, and integrated-circuits

2. Approach: concepts clustered around "integrated" or "integrated-circuits"

B

6) AELA Design of F8 x .2 IF-amplifier using integrated RC - network
7) AFSE Design and analysis of ac temperature compensated integrated amplifiers
8) AEAN Practical integrated-circuits of a typical stage of a cascaded video amplifier
9) ACOK Linear amplifier integrated-circuits
10) ABFX Research on compatible techniques for integrated-circuits
11) ADEY Design and fabrication of bandpass amplifiers using integrated-circuits
12) AFZS Optimizing stabilization-networks of functional components of integrated-circuits
13) ABTY Fabrication of unencapsulated hermetic-headers for integrated-circuits
14) ADMK Survey 1962 of microelectronics and integrated-circuits
15) ADSM Annotated bibliography 1963 of microminiaturization, thin-films and integrated-circuits
16) ADMG Terms and definitions used in microminiaturization and integrated-circuits
17) ACEM Component and circuit techniques for thin-film integrated-circuits and cryotrons
18) ACOH Design considerations for integrated-circuits
19) AFZT Packaging techniques for thin-film integrated-circuits, dot components, and micromodules

3. Approach: concepts clustered around "microelectronic" or "microelectronics"

see Figure 14b

Figure 14a

58 C

20) AECW Three-stage microelectronic dc-coupled amplifier with input-output feedback, gain -60
21) ADNG Deposition techniques for microelectronic modular interconnections
22) AFTL Relationship of power, gain, and stabilization in microelectronic small-signal amplifiers
23) AEDP Heat-transfer nomograph for evaluating thermal effects in microelectronic circuits
24) AFQE Development of microelectronic circuits with thin-film capacitors
25) ADWC Survey and prognosis of printed-circuits and microelectronics
26) AFNW Compromise in components and design tolerances for maximum reliability in microelectronics
27) ADMK Survey 1962 of microelectronics and integrated-circuits
28) AEQN History, techniques and progress 1962 in microelectronics and thin-films
29) ADNE Soviet advances in automation, microelectronics, and bioelectronics
30) AAAX Design criteria and evaluation of microelectronic systems
31) ADLW Design of microelectronic transistor video amplifiers
32) AFFZ Production details of microelectronic packaging techniques for transistors
33) ABCB Microelectronic reliability techniques for communication receiver design, F8
34) AEMY In-Sb infrared detector for temperature-profiling microelectronic-circuits
35) ADMK Survey 1962 of microelectronics and integrated-circuits
36) AFSO Microelectronic IF amplifier with thin-film passive components
37) ADLI Microelectronic tunnel-diode RF amplifier, F11 x .35, and video detector
38) ADRD Design and production of microelectronic diodes for micromodules
39) ADRC Design and production of microelectronic capacitors for micromodules
40) ADRF Design and production of microelectronic resistors for micromodules
41) ACQZ Survey of a thin-film microelectronic circuit
42) AFQE Development of microelectronic circuits with thin-film capacitors
43) ADNO Soviet article on printed resistors in microelectronic equipment

4. Approach: concepts clustered around "thin-film", or "thin-films"

see Figure 14c

Figure 14b

D

44) AFQE Development of microelectronic circuits with thin-film capacitors
45) AFDV Production techniques for multilayer subminiature thin-film ceramic capacitors
46) ABVD Tunnel-emission amplifier and its relation to thin-film circuits
47) AEPK TiO and SiO capacitors for thin-film circuits and Si-monolithic structures
48) AABY Study of miniature thin-film inductors 0.038-1.0mH, Q-100, F-G
49) AFZT Packaging techniques for thin-film integrated-circuits, dot components and micromodules
50) AEDU Analysis and fabrication of thin-film parametric amplifiers, F7 signal, F8-X.025 pump, gain 20
51) AFSO Microelectronic IF amplifier with thin-film passive components
52) AFBU Production and characteristics of thin-film pn-junction diodes and transistors
53) AFMT Characteristics of thin-film resistors and capacitors constructed of Ge-pn junctions
54) ABKK Parametric amplification in thin-film superconducting transmission lines
55) ABWW Theory of shot-noise in thin-film transistors
56) AEON Engineering problems associated with thin-film circuit manufacturing
57) AEMQ Resistance and capacitance limitations in multilayer thin-film circuit design
58) AGEA Thin-film fabrication techniques for conductors and resistors; PC-30 reproducibility
59) ACEM Component and circuit techniques for thin-film integrated circuits and cryotrons
60) ACGJ Production techniques for thin-film passive and active elements
61) ABDX Vacuum-deposition techniques for thin-film capacitors
62) ACUI Ta and Al thin-film devices showing amplification and energy-conversion properties
63) ACQZ Survey of a thin-film microelectronic circuit
64) ADET Preparation of thin-film capacitors by evaporation
65) ADSM Annotated bibliography 1963 of microminiaturization, thin-films and integrated-circuits
66) AEQN History, techniques and progress 1962 in microelectronics and thin-films
67) AAGH Survey of supercooled thin-film tunnel devices
68) AEVH Theoretical analysis of electron transmission coefficients in thin-film triodes
69) ABDX Vacuum-deposition techniques for thin-film capacitors
70) AATL Fabrication techniques for thin-film passive circuit components

Additional approaches by related Keywords that occur in concepts 1-70 (denoted by broken underlines where they appear first)

E

a) Molecular Electronics

71) AFZR Planar multiple diffusion method for fabricating functional blocks in molecular electronics
72) ABLG Airborne electronic-subsystems and circuits for molecular-electronics
73) AARL Differential amplifier using molecular electronics, common-mode-rejection 100db

59

Figure 14c

F b) Printed-circuits

76) ADAF Examples of printed-circuit-packaging
77) ADNO Soviet article on printed resistors in microelectronic equipment
78) ABBY Study of polyurethane coatings for printed-circuit assemblies

G c) Micromodules

79) ADRC Design and production of microelectronic capacitors for micromodules
80) ADRD Design and production of microelectronic diodes for micromodules
81) ADRF Design and production of microelectronic resistors for micromodules
82) AFZO Procedure for design and construction of micromodules

H d) Vacuum-deposition

83) AFVP High stability resistor produced by vacuum-deposition
84) AETP A masking method used in vacuum-deposition processes
85) ABDX Vacuum-deposition techniques for thin-film capacitors

Figure 14d

Question: Microminiaturization techniques for HF-amplifiers

alternative 1. approach: concepts clustered around "amplifier" or "amplifiers"

86) AFTL Relationship of power, gain, and stability in microelectronic small-signal amplifiers
87) AATD Low noise NF 2.7, low level tunnel-diode miniature amplifiers, F9
88) ABVD Tunnel-emission amplifiers and its relation to thin-film circuits
89) ADEY Design and fabrication of bandpass amplifiers using integrated-circuits
90) ABVB Molecular bandpass-and tunable amplifiers
91) ADLJ Design of microelectronic transistor video amplifier
92) ADLI Microelectronic tunnel-diode RF amplifier F11 x .35 and video detector
93) AECW Three Stage microelectronic direct-coupled amplifier with input-output feedback, gain 60
94) AELA Design of F8 x .2 IF amplifier using integrated RC network
95) ACOK Linear amplifier integrated-circuit
96) AGAN Practical integrated circuit of a typical stage of a cascaded video amplifier
97) ADEY Design and fabrication of bandpass amplifier using integrated circuits
98) AEDU Analysis and fabrication of thin-film parametric amplifier, F7 signal, F8 x .025 pump, gain 20
99) AFSO Microelectronic IF amplifier with thin-film passive components

Keywords with a broken underline may be used as guides to concept-clusters as listed on Figure 14a through 14c.

Figure 14e

Second-Generation Subject Organization
(Sample: Key word "Amplifiers")

A General

AO Reviews, Surveys, Bibliographics, Handbooks
AA Gain Control
AB Multistage
AC Noise Considerations
AD Dissipation
AE Distortion
AF Feedback in amplifiers
AG Capacitance in amplifiers
AH Miniaturization
AI Frequency response

B Theory and Design

BA Analysis and synthesis
BB Design
BC Stabilization
BD Research and development
BE Survey (see AO)

C Frequency

CA D.C.
CB Audio
CC R.F.
CD Microwave
CE Millimeter wave
CF Infrared
CG Optical

Figure 15

D Classification by Function

DA Intermediate and Carrier Frequency
DB Wideband and Narrowband
DC Linear
DD Differential and Difference
DE Pulse
DF Operational
DG Low-Noise
DH Decade
DI Phase-Inverter
DJ Low Distortion
DK Bipolar
DL Selective
DM Phase-Splitting
DN Frequency Converter
DO Band-, High-, Low- Pass

E Classification by Mode of Operation

EA Single-ended
EB Push-Pull
EC Feedback
ED Switching
EE Tuned
EF Pridge
EG Chopper
EH Parametric
EI Maser
EJ Laser
EK Negative
EL Cascade
EM Distributed
EN Traveling Wave

F Classification by Active Element

FA Diode (includes tunnel-diode, negative impedance)
FB Tube
FC Transistor
FD Magnetic Material

G Classification by Level

GA Low-Power
GB High-Power
GC Impedance
GD Voltage
GE Current

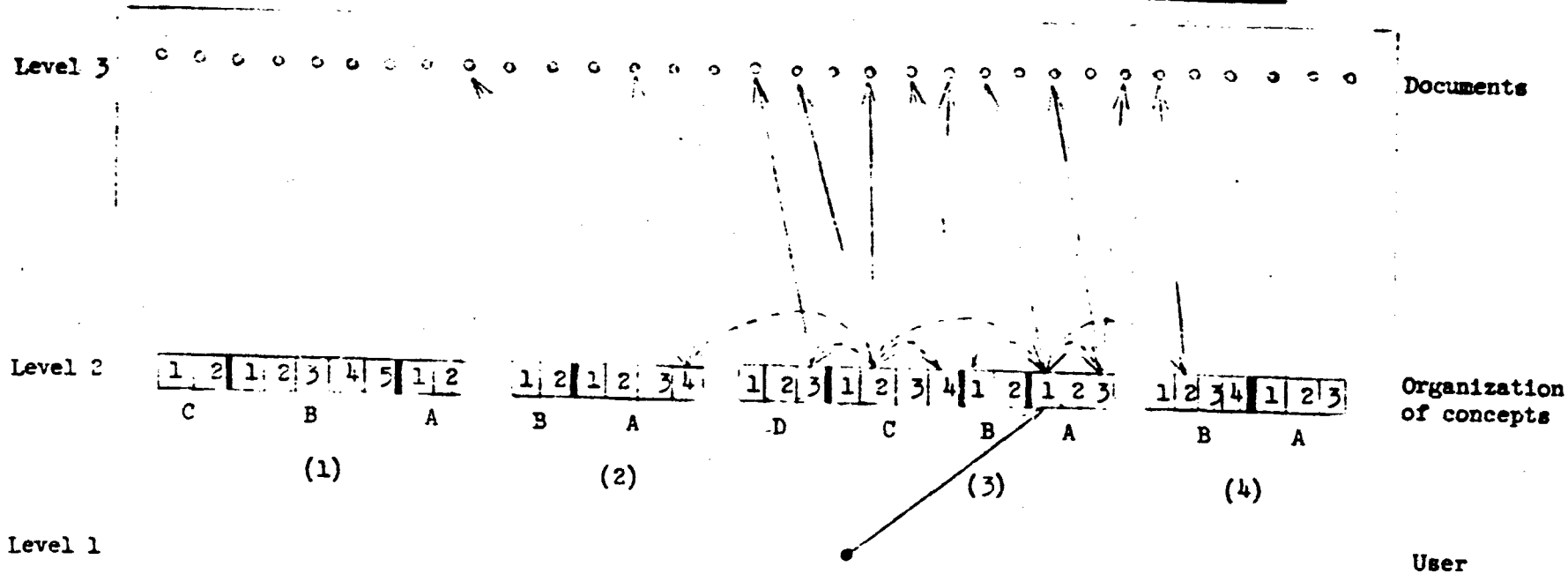
H Classification by Class of Operation

HA Class A
HB Class B
HC Class AB
HD Class C

I Classification by Specific Application

IA Radio
IB Television
IC Telephony
ID Telegraphy
IE Computers
IF Power
IG Control
IH Instrumentation
II Medical
IJ Optical
IK Industrial
IL Telemetry

Organization of Concepts in the Second-Generation Dictionary



Blocks (1) through (4) indicate families of concepts as arranged under frequent keywords like "laser," "amplifier," etc. Block (3) is divided into four groups according to four subheadings. Digits within these groups stand for individual concepts.

In this example, the user starts with keyword (3), where he finds the subheadings A, B, C, and D each consisting of several concepts. He selects the subheading A as the most suitable approach and starts his search with concept 1. Through this concept, he locates (a) one related document in the collection (indicated by a dark arrow) and (b) by using the keywords of this concept several additional concepts as indicated by dashed arrows (starting from concept 1). The additional concepts will also yield their corresponding documents as well as (through their constituent keywords pertinent to his question) "secondary" and "tertiary" approaches (indicated by dashed arrows) to further concepts.

Figure 15a

Retrieval using "ABC" and "Coordinate Indexing"

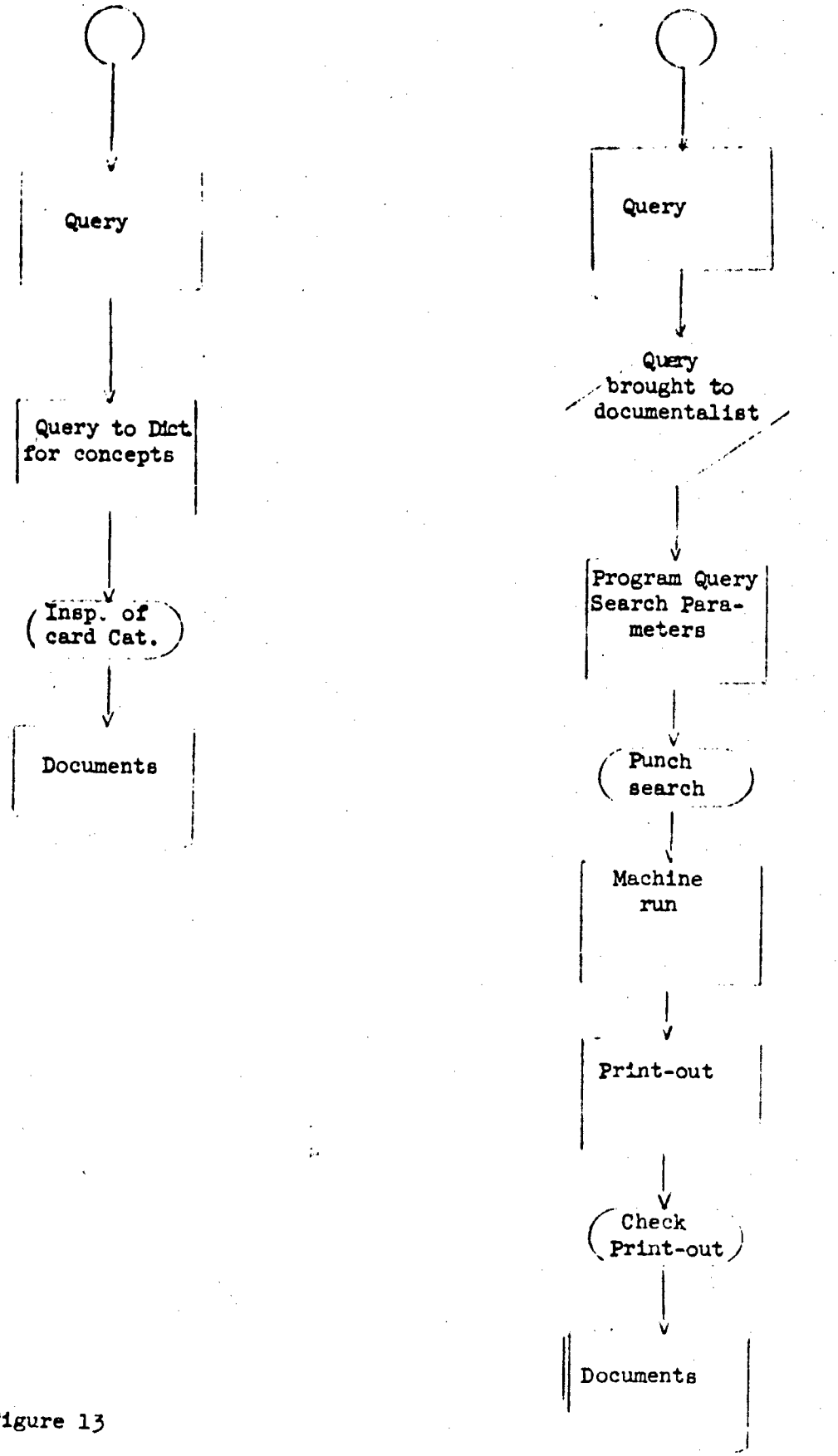


Figure 13

Test Operations by Groups

	Group I			Group II[d]
	Group 1[a]	Group 2[b]	Group 3[c]	
	A B			
I. Preparation of Questions:				
Type a: Document - Based	225			
Type b: Freely - Styled	36			
II. Standardization of Questions:				
Type a reduced to:				100
Type b:				36
III. Test Runs:				
Own Questions Type a:	50 50			
Counterpart's Questions Type a:	50 50	100	100	
Type b - Questions:	36 36	36	36	
TOTAL	136 136	136	136	
IV. Pre-evaluation of Results	136 136	136		
V. Final Evaluation of Results				136

[a] Consists of: 41 HDL scientists and engineers.
[b] Consists of: 6 Analysts (George Washington University).
[c] Consists of: 6 HDL Librarians.
[d] Consists of: ca 30 senior scientists and engineers including those of other agencies.

Chart I

FIRST-GENERATION ABC DICTIONARY (SAMPLE)

RUBY MASER•	AMPLIFIER	GAIN-14, F7-BANDWIDTH =	AEAY
DERIVATION OF POWER-GAIN IN TUNNEL-DIODE•	AMPLIFIER	=	AFGV
MINIATURIZED ELECTROCARDIOGRAM• CARRIER	AMPLIFIER	=	AFSI
EFFECT OF FET• IN LOW NOISE HIGH IMPEDANCE	AMPLIFIER	=	AGAU
ALYSIS OF PHASE DISTORTION• DUE TO AGC IN F8 TRANSISTOR	AMPLIFIER	= AN	AERU
-FREQUENCY COMPENSATION• OF DRIFT-TRANSISTOR AND OF TUBE	AMPLIFIER	= HIGH	ADKX
R, GAIN, AND STABILITY, IN MICROELECTRONIC• SMALL SIGNAL	AMPLIFIER	= RELATIONSHIP OF POWE	AFTL
TORTION• IN BALANCED AND UNBALANCED CLASS-AB PUSH-PULL	AMPLIFIER	= ANALYSIS OF HARMONIC DIS	AFIL
LITY• EQUATION FOR DIFFERENT TYPES OF TRANSISTOR VOLTAGE	AMPLIFIER	= DISCUSSION OF THERMAL STABI	AFWS
PLIFICATION AND GAIN OF FORWARD AND REFLECTED WAVE LASER•	AMPLIFIER	= THEORETICAL ANALYSIS OF AM	AEUG
EL-DIODES• = NEGATIVE-RESISTANCE	AMPLIFIER	AND BISTABEL PULSE-CIRCUITS USING TUNN	AFCH
PERFORMANCE OF PARAMETRIC-DIODES• IN	AMPLIFIER	AND FREQUENCY MULTIPLIER-CIRCUITS =	AFOJ
TUNNEL-DIODE• EQUIVALENT-CIRCUIT,	AMPLIFIER	AND FREQUENCY CONVERTER =	AFDT
INSTABILITY IN THE NEMAG•	AMPLIFIER	AND GENERATOR =	ADZW
DESIGN AND ANALYSIS OF F9 NEGATIVE-RESISTANCE•	AMPLIFIER	AND OSCILLATOR USING TUNNEL-DIODE• =	ABHF
HIGH-FREQUENCY-TRANSISTOR F9, PN2 AS OSCILLATOR OR	AMPLIFIER	AT TWICE THE CUTOFF-FREQUENCY =	AAWS
FREE-RUNNING AND PULSE	AMPLIFIER	AVALANCHE SWITCH-CIRCUITS• =	AFWX
ENT OF TRANSISTOR• PARAMETERS AND EQUIVALENT-CIRCUIT FOR	AMPLIFIER	DESIGN = MEASUREM	ADLV
PUT = CONSTANT TEMPERATURE ONE-TRANSISTOR	AMPLIFIER	FOR SE PHOTOVOLTAIC-CELLS• VP2X.6-OUT	AEMK
POWER-SUPPLY, CHOPPING OSCILLATOR AND TUNED	AMPLIFIER	FOR SENSITIVE PHOTOMULTIPLIER• =	AEQY
OMETER• = HYBRID DC	AMPLIFIER	FOR THERMOCOUPLE• AND RESISTANCE THERM	AFFB
GE OF 80-DB = LIMITER•	AMPLIFIER	OPERATING AT F7X.8 WITH A DYNAMIC-RAN	AGFW
BIAS OF TUNNEL-DIODE• USED AS SWITCH OR	AMPLIFIER	OR OSCILLATOR =	AAXS
HING-EQUIPMENT• = TRANSISTOR FERRITE-CORE	AMPLIFIER	PROVIDES BASIC LOGIC-CIRCUIT FOR SWITC	AFWE
MINIATURIZED ELECTROCARDIOGRAM• TRANSISTORIZED CARRIER	AMPLIFIER	SYSTEM =	AFVM
HALL-EFFECT MULTIPLIER• USING FEEDBACK	AMPLIFIER	TO REDUCE PHASE-SHIFT-DISTORTION =	AAVL
INDUSTRIAL PREPAREDNESS-STUDY• FOR DEVICE 7, F8X.7	AMPLIFIER	TRANSISTOR=	AFFY
PHASE-SENSITIVE 5-STAGE TRANSISTOR HOMODYNE-TYPE AC	AMPLIFIER	USED IN RADIATION DETECTOR• =	AFHX
= EXCITATION OF 8-INCH RUBY LASER•	AMPLIFIER	USING ELLIPTICLA XE FLASH- LAMPS T300	ADWX
= DESIGN ANALYSIS OF HIGH INPUT IMPEDANCE	AMPLIFIER	USING FET• AND NPN BIPOLAR-TRANSISTOR	AGDM
STABLE DC OPERATING-POINT OF TRANSISTOR	AMPLIFIER	USING ZENER-DIODES =	AFAY
PEDANCE• INDEPENDENT OF OUTPUT-LOAD = FEEDBACK	AMPLIFIER	WITH ADJUSTABLE TO INFINITY INPUT-IM	AAUQ
TEMPERATURE-STABILIZATION• OF TRANSISTOR	AMPLIFIER	WITH ZENER-DIODE FEEDBACK-NETWORK =	AFBS
WIDE-FREQUENCY RANGE GAIN-CONTROL• FOR TRANSISTOR	AMPLIFIER	WITHOUT VARYING DC BIAS =	AGDL
DESIGN OF A COMMON-EMITTER DISTRIBUTED	AMPLIFIER•	FO TO F9 =	ADFI
DESIGN OF RESISTANCE-COUPLED TRANSISTOR	AMPLIFIER•	F4-5, GAIN-P60 =	ADKS
TRANSISTOR RC-COUPLED HIGH-FREQUENCY	AMPLIFIER•	F8, GAIN-P12 =	AEDH
ONE-STAGE HIGH-GAIN TRANSISTOR	AMPLIFIER•	F8, GAIN-P8.5 =	AEDG
THREE-STAGE CASCADED	AMPLIFIER•	F8X, F8X.32-BANDWIDTH, GAIN-P34 =	AEDI

Chart II

• • • AMPLIFIER -- GENERAL • • •

• • • AMPLIFIER -- GAIN CONTROL • • •

ALYSIS OF PHASE DISTORTION• DUE TO AGC IN FB TRANSISTOR AMPLIFIER • AM AERU
WIDE-FREQUENCY RANGE GAIN-CONTROL• FOR TRANSISTOR AMPLIVIER WITHOUT VARYING DC BIAS • AGOL

• • • AMPLIFIER -- MULTISTAGE • • •

NOMOGRAPH FOR DESIGN OF RC-COUPLED TRANSISTOR AMPLIFIER• • ACQT
FREQUENCY-RESPONSE OF TRANSISTOR TRANSFORMER-COUPLED AF AMPLIFIER• • APMO
MATRIX OF ONE-ARM OF WIEN-BRIDGE APPLIED AS SELECTIVE RC AMPLIFIER• • DERIVATION OF THE ADGC
DESIGN OF RESISTANCE-COUPLED TRANSISTOR AMPLIFIER• P4-9, GAIN-P60 • ADAS
TRANSISTOR RC-COUPLED HIGH-FREQUENCY AMPLIFIER• FB, GAIN-P12 • ABDM
THREE-STAGE CASCADED AMPLIFIER• FBX, FBX.32-BANDWIDTH, GAIN-P34 • AEDI
MAXIMUM GAIN BANDWIDTH PRODUCT FOR 2-TERMINAL TRANSISTOR AMPLIFIER• INTERSTAGE • APAN
OF TRANSISTORIZED DISTRIBUTED- AND CASCADED- AMPLIFIER• AMPLIFIER• •/SIS AND GAIN-BANDWIDTH COMPARISON ACOL

• • • AMPLIFIER -- DISTORTION • • •

ALYSIS OF PHASE DISTORTION• DUE TO AGC IN FB TRANSISTOR AMPLIFIER • AM AERU
PREDISTORTION METHOD OF MULTI- TUNNEL-DIODE-AMPLIFIER• AMPLIFIER• DESIGN • ABAC

• • • AMPLIFIER -- FEEDBACK IN AMPLIFIERS • • •

HALL-EFFECT MULTIPLIER• USING FEEDBACK AMPLIFIER TO REDUCE PHASE-SHIFT-DISTORTION • AAVL
TEMPERATURE-STABILIZATION• OF TRANSISTOR AMPLIFIER WITH ZENER-DIODE FEEDBACK-NETWORK • AFBS

• • • AMPLIFIER -- MINIATURIZATION • • •

R, GAIN, AND STABILITY, IN MICROELECTRONIC• SMALL SIGNAL AMPLIFIER • RELATIONSHIP OF POWE APTL
DESIGN AND FABRICATION OF BANDPASS AMPLIFIER• USING INTEGRATED-CIRCUITS • ADEY

• • • AMPLIFIER -- THEORY AND DESIGN • • •

• • • AMPLIFIER -- ANALYSIS AND SYNTHESIS • • •

ANALYSIS AND DESIGN OF P2X.25 COMPENSATED TRANSISTOR AMPLIFIER• • APCE
MPARISON, AND DESIGN OF BROAD-BANDED, LOW-PASS TRANSISTOR AMPLIFIER• • ANALYSIS, CO APMF
GENERAL SYNTHESIS OF TUNNEL-DIODE AMPLIFIER• AND SENSITIVITY-MINIMIZATION • APEW

• • • AMPLIFIER -- DESIGN • • •

ANALYSIS AND DESIGN OF P2X.25 COMPENSATED TRANSISTOR AMPLIFIER• • APCE
DESIGN ANALYSIS OF SINGLE-TUNED TRANSISTOR BANDPASS AMPLIFIER• • APMI
HO GRAPHIC DESIGN FOR SERIES TUNED NEGATIVE-RESISTANCE• AMPLIFIER• • ANALYTIC A AEMC
MPARISON, AND DESIGN OF BROAD-BANDED, LOW-PASS TRANSISTOR AMPLIFIER• • ANALYSIS, CO APMF

• • • AMPLIFIER -- APPLICATIONS • • •

MINIATURIZED ELECTROCARDIOGRAM• CARRIER AMPLIFIER • APSI
AN EXPERIMENTAL TUNNEL-DIODE AMPLIFIER• FOR P7-8 APPLICATION • APQY
TRANSISTOR-RELAY AMPLIFIER• FOR ANALOG-COMPUTER • ACDC
FREQUENCY-CONVERSION MAGNETIC REPEATER AMPLIFIER• FOR LOW-LEVEL SUBMARINE TELEGRAPH-CABLES ACBT
CONSTANT TEMPERATURE ONE-TRANSISTOR AMPLIFIER FOR SELENIUM PHOTOVOLTAIC-CELLS• VP2X.6-OUTPUT • AEMR
POWER-SUPPLY, CHOPPING OSCILLATOR AND TUNED AMPLIFIER FOR SENSITIVE PHOTOMULTIPLIER• • AEQV
HYBRID DC AMPLIFIER FOR THERMOCOUPLE• AND RESISTANCE THERMOMETER, ALSO FOR OTHER PURPOSES • APFB
TRANSISTOR FERRITE-CORE AMPLIFIER PROVIDES BASIC LOGIC-CIRCUIT FOR SWITCHING-EQUIPMENT• • AFME
A DETAILED DESCRIPTION OF THE DESIGN OF A MINIATURIZED ELECTROCARDIOGRAM• TRANSISTORIZED CARRIER AMPLIFIER SYSTEM • APUN
SIX-CHANNEL DC-COUPLED AMPLIFIER• USED AS GALVANOMETER-DRIVER UP TO P4 • ACOX
STABLE TRANSISTOR AMPLITUDE COMPRESSION AMPLIFIER• USED AS MULTICHANNEL DISTRIBUTOR • APEI
EXPERIMENTAL EVALUATION OF THE PHASE-SENSITIVE 5-STAGE TRANSISTOR HOMODYNE-TYPE AC AMPLIFIER USED IN THE HOLMES-JOHNSON AND LAROCHE RADIATION DETECTORS• • APMK

• • • AMPLIFIER -- CHARACTERISTICS • • •

EFFECT OF FET• ON LOW NOISE HIGH IMPEDANCE AMPLIFIER • AGAU
RELATIONSHIP OF POWER, GAIN, AND STABILITY, IN MICROELECTRONIC• SMALL SIGNAL AMPLIFIER • APTL
A BASIS FOR THE ANALYSIS OF HARMONIC DISTORTION• IN BALANCED AND UNBALANCED CLASS-AB P~-PULL AMPLIFIER • APIL
ANALYSIS, COMPARISON, AND DESIGN OF BROAD-BANDED, LOW-PASS TRANSISTOR AMPLIFIER• • APMF
SOME COMMENTS ON THE INFLUENCE OF STABILITY ON NOISE-FIGURE FOR NEGATIVE-CONDUCTANCE AMPLIFIER• • ABZZ
DESIGN CONFIGURATION AND PARAMETERS OF NEGATIVE FEEDBACK AMPLIFIER WITH ADJUSTABLE TO INFINITY INPUT-IMPEDANCE• INDEPENDENT OF OUTPUT LOAD AND CERTAIN OTHER PARAMETERS • AAUZ
TRANSISTOR AMPLIFIER• ZP6-INPUT F-1-5-BANDWIDTH FOR HIGH-IMPEDANCE MICROPHONE • AAPB

Chart III

Sample of Parameters (First-Generation)

FREQUENCY	(in cps)	
Fa		means 10^{a-1} to 10^{a} cps
Example:	F2	means 10^{1} to 10^{2} cps
FaX.b		means .b X 10^{a} cps
Example:	F5X.3	means .3 X 10^{5} cps
FaX.b-c		means .b X 10^{a} to c X 10^{a} cps
Example:	F2X.5-6	means .5 X 10^{2} to 6 X 10^{2} cps
FaX.b-cX.d		means .b X 10^{a} to .d X 10^{c} cps
Example:	F2X.3-5X.4	means .3 X 10^{2} to .4 X 10^{5} cps

CURRENT	(in amperes)	
APa		means 10^{a-1} to 10^{a} amps
Example:	AP3	means 10^{2} to 10^{3} amps
ANA		means 10^{-a} to 10^{-a+1} amps
Example:	AN5	means 10^{-5} to 10^{-4} amps
APaX.b		means .b X 10^{a} amps
Example:	AP2X.3	means .3 X 10^{2} amps
ANaX.b		means .b X 10^{-a} amps
Example:	AN3X.4	means .4 X 10^{-3} amps
APaX.b-c		means .b X 10^{a} to c X10^{a} amps
Example:	AP2X.3-7	means .3 X 10^{2} to 7 X 10^{2}amps

Chart IV

Supplement

Computer Programs of the HDL Information System

by

William G. Brown

I. Present System

The over-all planned system for library automation is shown in Figure A. The function of each of the functional blocks of the planned system is as indicated. However, at the present time, only four of the indicated functions are operational: descriptive cataloging, subject analysis, ABC dictionary updating, and subject card catalog updating. In more general terms, these four functions are conveniently grouped together as only two broad functions, namely, cataloging and ABC dictionary updating.

It should be emphasized that the presently operating versions of these systems for cataloging and ABC dictionary updating are not the same as those planned for implementation in the future. Several refinements for more simple operation as well as for more attractive outputs will be incorporated into the final system. However, before proceeding to a description of the final system, those portions of the systems presently operating will be described.

Flow diagrams of the two presently operating systems are shown in Figures B and C. It should be noted that each rectangular block in these flow diagrams indicates a separate computer program. The circular symbols are tapes, with the drive numbers on which they are mounted indicated as "DR x" where appropriate.

A. Cataloging System

At present the cataloging system produces two-part accession bulletins, catalog cards, and appropriate files of information on magnetic tape. The two-part accession bulletins are composed of: a) a bibliographic listing (in broad-subject-category order) printed by the Bulletin Print program, and b) a KWIC rotated title list prepared from a tape output of the Bulletin Print program by the BE-PIP (Bell Permutation Index Program) 7090 program, supplied by the IBM SHARE system from the original author, the Bell Laboratories. The catalog cards are printed from a tape prepared by the Bulletin Print program only after this tape has been sorted into approximate filing order for ease of placing the cards in the catalog drawers. The tape files maintained are useful for such operations as subject-card-catalog updating, and the possible printing of additional catalogs and lists (such as the recently provided lists of corporate authors and contract and project numbers). The operation of the cataloging system (Figure B) is as follows:

1. As input to the cataloging system, IBM cards are punched in a format designed especially for this application. These cards are punched from a worksheet (Figure D) on which each line represents a single IBM punched card. In this format, the shelf number of each item (card columns 1 -- 8) is repeated on each punched card, as is the Grp/Seq (Broad-Subject Category) number

(Column 74 -- 79).* Thus, the cards to be cataloged for a given item are easily sorted together on these numbers. However, in order to: a) identify what portion of an entry is signified by a given card, and b) properly sequence the cards within a given type of card (say a title card), two additional numbers are added to each punched card. These are the card numbers contained in Columns 10 and 11. The first of these (Column 10) indicates the type card, according to the type of information entered into it (report number, corporate author, contract/project number, title, personal author, subjects, etc.). The second digit (Column 11) merely provides for proper sequencing of cards within a similar type (Column 10) entry. The data which are subsequently used for printing are all (except for shelf number) punched in Columns 12 - 62.

2. IBM punched cards are put onto tape with the TFG-B program, which is a utility program supplied by IBM.

3. The tape images of the cards are then sorted into order by the Sort/Merge 11 Program supplied by IBM. Sort/Merge 11 is used in making all sorts and merges indicated. The order into which the card images are sorted is the following:

   a. Subject-Category number,
   b. Shelf number, and
   c. Card sequence number within each cataloged item.

4. The card images are then run through a purge program, which deletes any items containing detectable errors. These items may then be corrected and re-introduced into the system in the next run.

5. The records which are not deleted by the Purge Program are then used for two purposes. In one instance they are merged with the input file (on tape), which is a file of all records which have previously been entered into the system. The other use of the records is as input to the Bulletin Print Program.

6. The Bulletin Print program prepares several output tapes while it is printing a bibliographic listing of all valid items entered into it. One of these output tapes, DR 1, is sorted for subsequent use in printing catalog cards. Another tape, DR 4, is used as input to the Bell BE-PIP program used to produce a rotated title list. Still a third tape, DR 5, is used to maintain a Partial Subject File for subsequent use in updating the subject card catalog with the ABC dictionary updating system.

7. The sorting of the Drive 1 output tape (the main cataloging output) is done in the following way:

   a. Card type (shelf number card, Project number card, report number card, subject card, etc.)
   b. Data entered onto the card sorted upon.

---

*Only the first three positions are currently used. These represent a broad subject category, which is used at present only to organize the periodical accessions bulletins into subject categories. Thus, the primary sorting of inputs is by this subject category, the secondary sorting is by the shelf number, and the tertiary sorting is by the two-digit card numbers (Columns 10 and 11).

In other words, the tape file is sorted first upon the type of card catalog entry to be made, and second, upon the alpha-numeric information of that particular entry.

8. The Card Print program has only to print the sorted output from Drive 1 of the Bulletin Print program to create a set of catalog cards that are in approximate filing order. The sorted tape is also used to add to the tape file of all cataloging information.

9. The Drive 4 output tape from the Bulletin Print program is run through a series of programs to produce the desired outputs. The first such program is the previously mentioned BE-PIP program. Although this program produces several output (tape) files, only one is utilized for the cataloging system. This is the second file contained on Drive A-5. This tape is entered into a 1410 program which does two things: a) it prints the rotated title listing, and b) it puts the file out on another tape, as the first file on that tape. This output (first-file) tape is then merged with the file of all rotated titles thus far entered into the system. This longer file may be used to occasionally print rotated lists over longer periods.

B. ABC Dictionary Updating System

The ABC Dictionary Updating System (Figure C) must perform several functions. To begin with, it must provide for making additions and changes to the entries in the dictionary. The changes are needed in order to further standardize terminology and asterisk terms. The additions, of course, will always be needed in any open-ended system. But another function, that of making deletions, also is needed. Although this would appear simplest of all, there is an additional requirement that, since any deletion is made only in order to combine entries having similar meanings, the reports cataloged under the deleted entry must automatically be transferred to the other entry having the same meaning. The operation of the system is as follows:

1. The old version of the dictionary input tape and a tape of the desired changes are entered into the ABC Dictionary Update program. The types of changes are additions, changes to current entries, and deletions with replacements. The program produces three output tapes: a) a new updated version of the dictionary input tape, b) a list of deletions and their replacments (on tape), and c) a list (on tape) of all valid asterisk terms for each coded item in the new updated version of the dictionary.

2. The new dictionary input tape is then entered into the BE-PIP program (the same one used in the Cataloging system) so as to produce a list (on tape) of rotated concepts. Also as in the cataloging system, the BE-PIP Drive A-5 output tape is printed and made into a first-file tape, which can be used to print additional copies of the dictionary more rapidly. This is all that is required to update the dictionary itself.

3. Both of the other tapes from the ABC Dictionary Update program are sorted into code order (separately), for entry into additional programs.

4. The Drive 4 output from the Dictionary Update program (sorted) is entered into the Delete and Replace Asterisk Terms program, along with the Drive 5 output from the Dictionary Update program (sorted). The Delete and Replace Asterisk Terms program then inserts the proper asterisk terms for the coded items which are replacing the deleted items, and the results are written out on the Drive 3 tape as changes.

5. The Drive 5 output from the Dictionary Update program is entered (sorted) into the Compare and Change Asterisk Terms program, along with the Drive 5 output from the previous updating run. The Compare and Change Asterisk Terms program compares each set of asterisk terms for each coded item, and whenever the new asterisk terms are different from those previously used, it writes these new ones out as changes on Drive 3.

6. The two Drive 3 output tapes from the Delete and Replace Asterisk Terms program and the Compare and Change Asterisk Terms program are then sorted together, and are entered on Drive 2 into the Change Reports Subject File program.

7. The change records on tape Drive 2 are compared with the Reports Subject File records on Drive 1 by the Change Reports Subject File program. This program then produces three output tapes. One of these tapes is a new updated version of the Reports Subject File, containing all changes caused by the updating of the dictionary; this is the Drive 3 output. Another output is that on Drive 4, which is the same as the Drive 3 output, but which contains only the items to which changes have been made, rather than the entire Reports Subject File. The third output is that on Drive 5, which is a list of any changes for which reports were not found.

8. The Drive 4 output is then sorted on its asterisk term and code, and is put into the normal Card Print program to produce additional entries to the subject card catalog. It should also be noted that, although it is not shown in the diagram, the Drive 3 output file must also be re-sorted before being used again as the Drive 1 input tape.

II. Planned System

The final system presently in the development stages will be a substantially modified version of the presently operating system. In general, the final system will incorporate such changes as: a) somewhat expanded main cataloging tape record to allow for insertion of codes to tie together similar bits of information that are parts of the same entry; b) a two-level sorting field in the main cataloging tape record to allow for sorting of catalog cards not only by the primary data (such as the corporate author) but also by secondary data (such as a report number or project number); c) an automatic procedure for calling in various programs as needed to process those transactions for which entries have been made (such as cataloging, charge or discharge, request for purchase, etc.); and d) the ability to process records for all types of materials currently held by the HDL library (books, periodicals, reports, proceedings, etc.) rather than merely for technical reports. The operating

characteristics of this planned system are approximately as follows:

1. As in the present system, IBM cards will be keypunched from a specially formatted work-sheet. However, in order to allow for additional needs of the planned system, such as the longer shelf numbers used for books, the key punching format is somewhat different. The planned work-sheet is shown in Figure E. On this work-sheet there are two card columns (1 and 2) which indicate the nature of the transaction being processed (cataloging entry, purchase order, charge/discharge, updating of previously cataloged material, etc.) and the type of document to which the transaction applies (periodical, book, technical report, bound periodical volume, etc.). Another added feature is the incorporation of a Change (CHG) column, which permits the correction or deletion of previously punched cards by the method of simply replacing or deleting them with another card, rather than locating them in the deck. Similarly, the card being punched may also be rendered invalid by putting the proper punch in the CHG column. There is also provision for using the CODE columns on all cards, so as to be able to tie together items which match logically.

2. The punched cards are put onto tape by a special Card-to-Tape program, rather than the IBM-provided utility program. This special Card-to-Tape program is needed to provide for changing the Transaction and Document codes from letters which are somewhat mnemonic (a Book is a "B", a periodical is a "P", a Chargeout is a "C", etc.) to letters which provide the proper sequence of internal machine operations when sorted upon. Thus, the human key-punching problem is made simpler, while the order of operations in the machine system is allowed to be optimum.

3. After the cards are put onto tape in 80-column format, they are sorted as follows:

   a. Transaction Code,
   b. Document Code,
   c. Category (Subject category),
   d. Shelf Number,
   e. Card Number, and
   f. Change Number.

This places all input cards on the tape in the proper sequence, so that those cards affecting earlier operations, such as pre-cataloging and cataloging, may be processed and entered into the master tape files before those cards affecting necessarily later operations, such as charge/discharge, are to be processed. In addition, this program will provide some reformatting needed for internal operations.

4. All of the programs needed to process the input cards will be stored on a single magnetic tape, and the system will proceed from one operation to the next, in order, automatically. The only operator intervention needed will be that of mounting the proper tapes and paper forms on the peripheral equipment when called for by the programs. It should be pointed out, however, that at least two versions of the system will be available, one for daily runs and another for weekly runs. There will also be a third option for special runs to be run only once every three months.

5. Cards which are not processed during a given run will be stored separately on another tape so as to be available for subsequent runs in which they may be processed.

6. The acquisitions sub-system of the final design allow for both the purchasing and the requesting of free copies of all library materials desired. This system will be so designed as to provide a pre-cataloging tape record at the time of the request or order, all or part of which can be used as the final cataloging record, thereby eliminating a portion of the typing and key-punching need later. In addition, the system will have the capability of providing budgetary assistance to the Librarian who selects items for purchase. When items are selected, they will be assigned a single digit priority (1 -- 9), or a special code for immediate purchase. Then, when the purchasing run is made, the available funding will be inserted from the previous run, updated if necessary, and the program will automatically select all items beginning with the highest priority (1) until the funds available for the period have been expended. These will then be sorted by sources, and purchase request forms will be printed. Those items which are not purchased will be kept on a tape for possible purchase during some future run, either as the result of normal operations or as the result of increased funding. Items for which orders are subsequently cancelled will result in the available funding figure being appropriately increased.

7. The Cataloging process will provide much the same outputs as that of the original system. However, the 3 x 5 inch catalog cards will be printed out in a two-level filing order, rather than the single level order now provided. In addition, some refinements will be made so as to allow the coupling of additional corporate authors to additional project or report numbers, and any other such couplings as may be needed. The new tape formats will also allow for easier file maintenance operations (on tape).

8. The ABC Dictionary Update system, while much like that presently used will have at least one advantage over the present system. That is, whereas the present system requires the complete running of the KWIC program for each update (about 20 minutes on the 7094), the final version will require the running of only those items for which changes, additions, or deletions have been made.

9. The automatic Dissemination system will make it possible to disseminate materials on the basis of four criteria — project or contract number, ABC concept, broad subject category, or special document tracings. Thus, it will be possible for scientists who have a continuing interest in a particular project or concept to be notified of the availability of newly received information pertinent to that interest. In addition, special projects having a need for all available information in a particular broad field may be notified of its presence in the system.

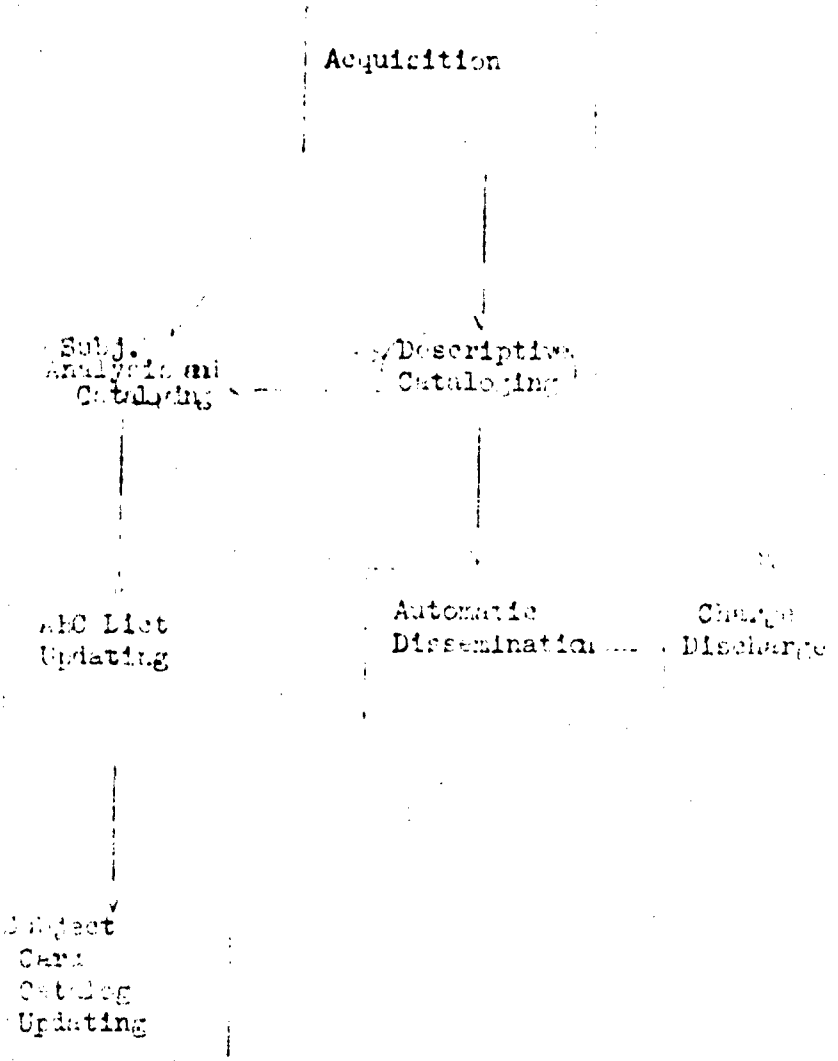
10. The Charge/Discharge system will perform all of the functions normally associated with the automatic maintenance of chargeout information. However, in the HDL environment, it is necessary that the processing of these records allow for the rapid location of any item that is charged-out. The system will provide for this need, and will, in addition, on request, print out a shelf-list of all items

that should be on the shelf at any given time, to allow for ease of making a shelf reading (inventory). The system will also, of course, provide automatically for the printing out of pre-addressed over-due notices (address information will be taken at the time of recall from a master personnel-office tape record). The system will also provide lists of items held by individual borrowers, either on request, or automatically in the event that the personnel tape record indicates the impending leaving of an individual. The system will permit the automatic performance of all transactions by means of a Frieden Collectadata 30 System or a similar equipment accepting the identification number of the borrower punched on his charge plate together with the pre-prepared catalog information (a punched card produced from a tape).

Over-all, the final system will provide for easier operation, both by library personnel and by computer operators. The simplification for library personnel results largely from the provision that all cards for all types of transactions can be entered into the system without regard for batching or sorting. In addition, the final system formats were created with a view to possibly utilizing paper-tape typewriters or even magnetic-tape typewriters, if such should ever be deemed feasible. Further, the filing of catalog cards by two levels will be as nearly automatic as possible. The ease of operation for the computer operators is obvious, since they will no longer be called upon to select a sequence of several programs and run them in turn, but will merely be called upon to mount tapes and so forth as called for by the system at the time of operation.

The final system will thus provide for a far better man-machine relationship than the original system, and should assist the library significantly in its operations.

Over-All Planned System



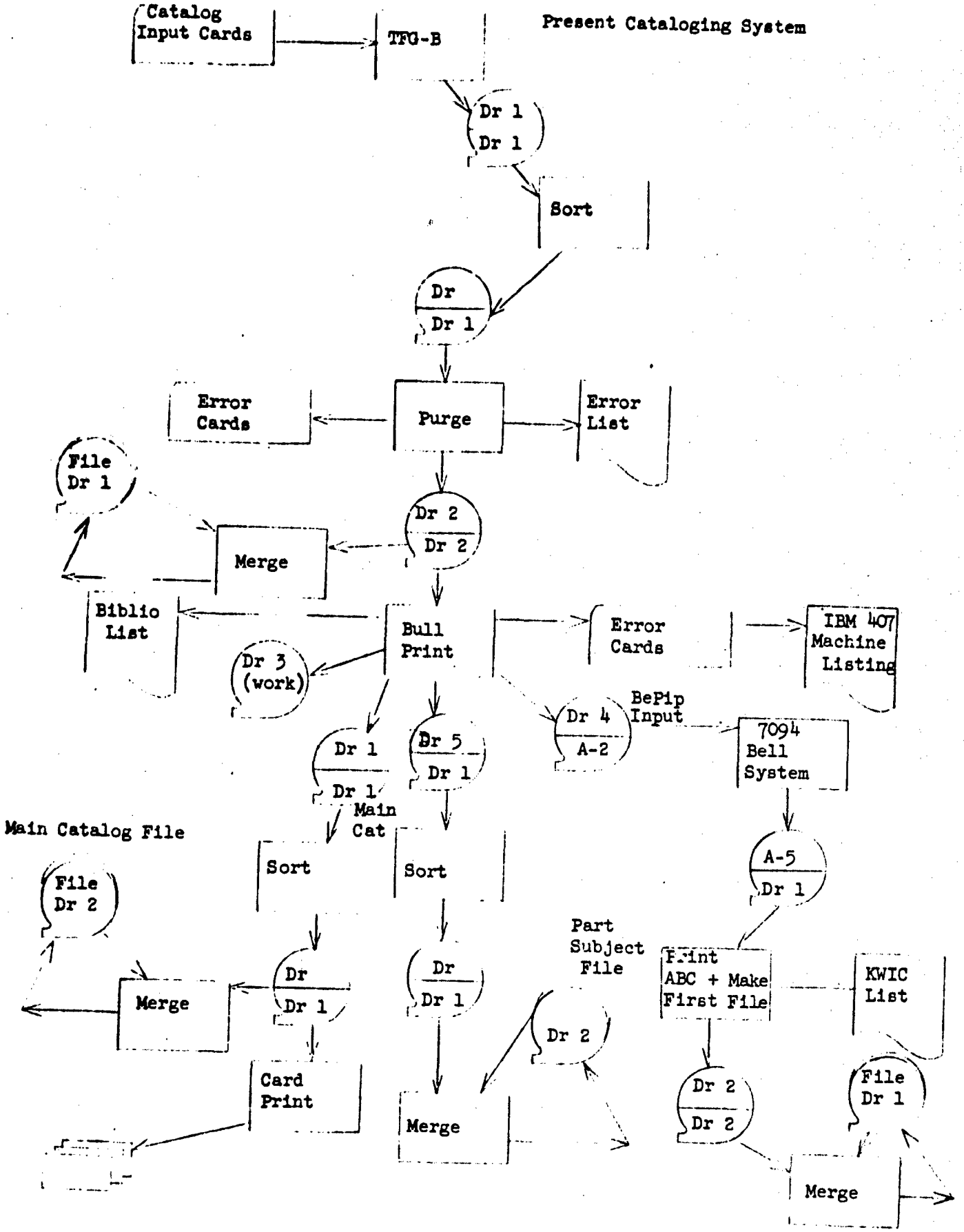
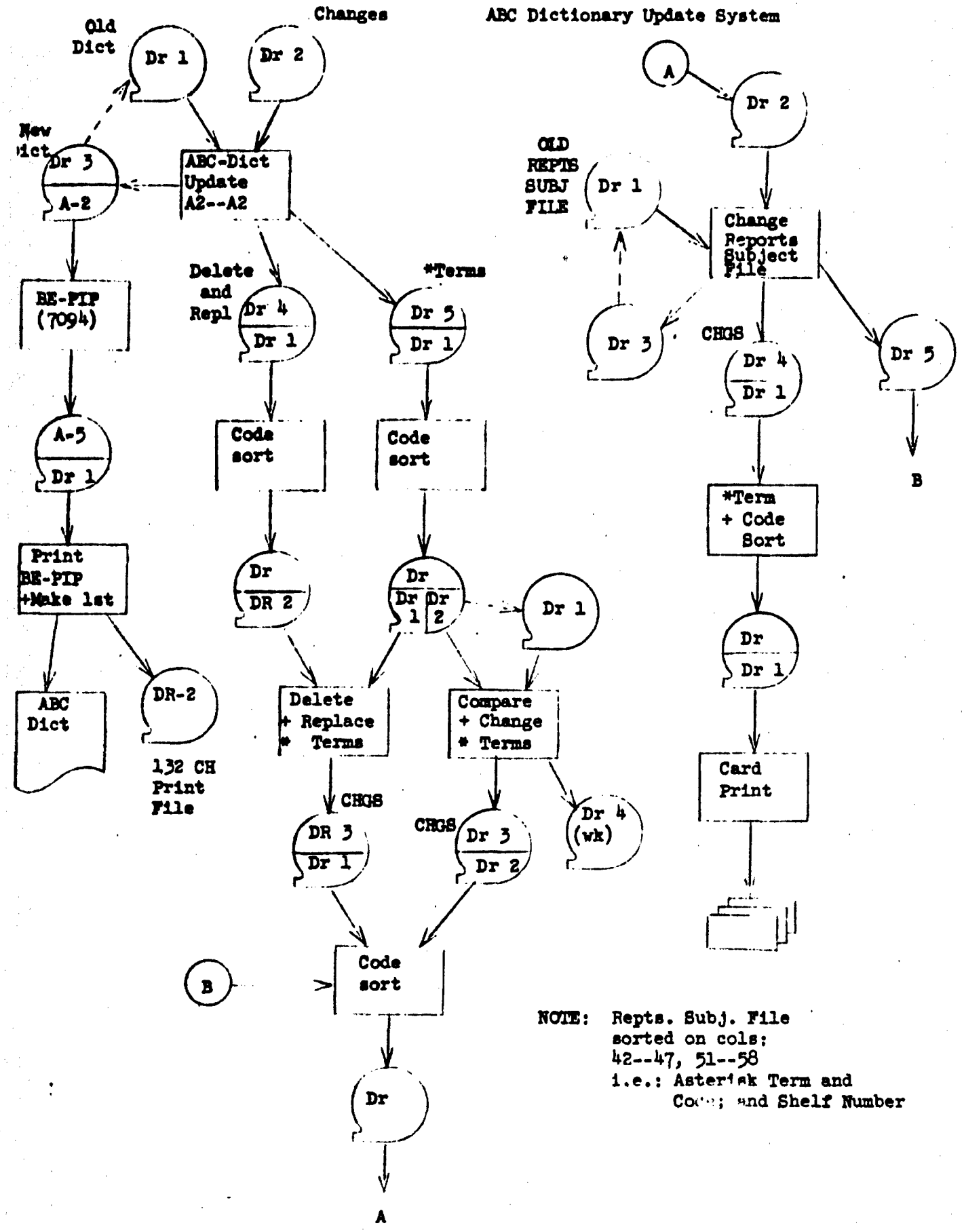

Present Cataloging System
Catalog Input Cards
TFG-B
Dr 1
Dr 1
Sort
Dr
Dr 1
Error Cards
Purge
Error List
File Dr 1
Dr 2
Dr 2
Merge
Biblio List
Bull Print
Error Cards
IBM 407 Machine Listing
Dr 3 (work)
Dr 4
A-2
BePip Input
7094 Bell System
Dr 1
Dr 1
Dr 5
Dr 1
Main Cat
Main Catalog File
Sort
Sort
A-5
Dr 1
File Dr 2
Part Subject File
Print ABC + Make First File
KWIC List
Dr
Dr 1
Dr
Dr 1
Merge
Dr 2
Card Print
Merge
Dr 2
Dr 2
File Dr 1
Merge

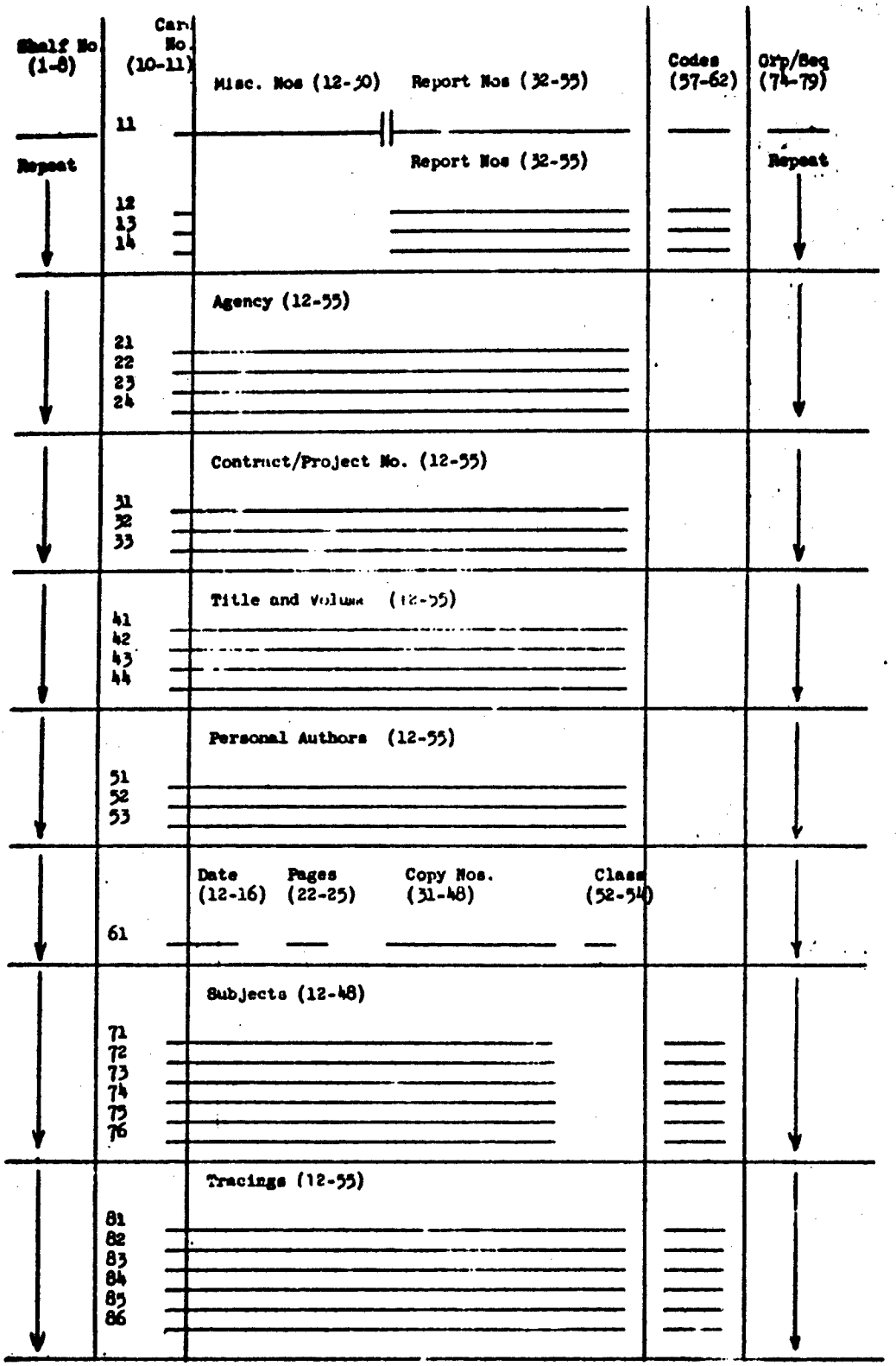
Figure B



NOTE: Repts. Subj. File sorted on cols: 42--47, 51--58 i.e.: Asterisk Term and Code; and Shelf Number

Figure C

ABC-Dict Update System (2-16-65)



AMCDO FORM 476
(Rev) 9 July 1963

LIBRARY CATALOGING WORKSHEET

This form supersedes ORDTL 476 dated 1 Dec 1961

Figure D

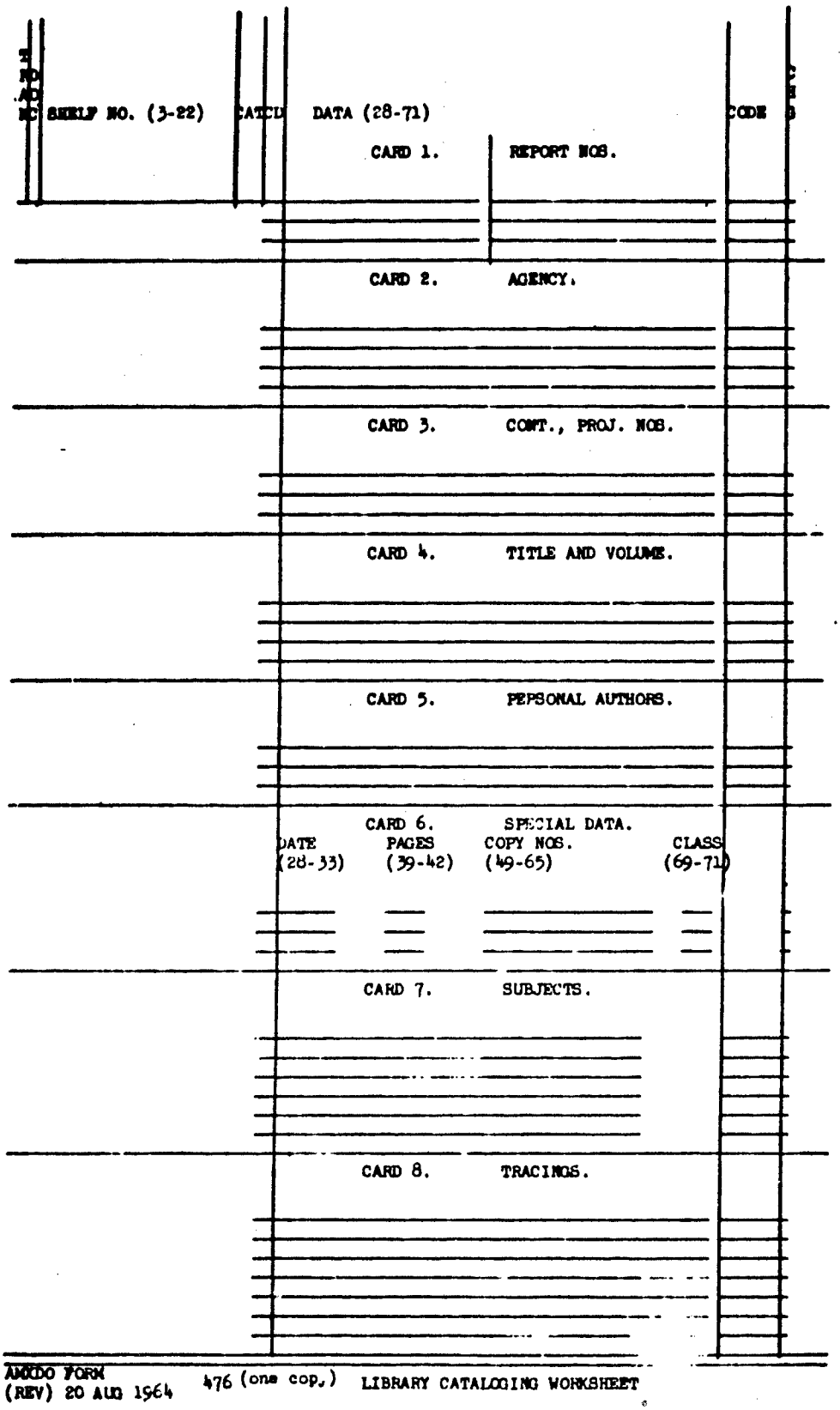

SHELF NO. (3-22)
DATA (28-71)
CODE
CARD 1. REPORT NOS.
CARD 2. AGENCY.
CARD 3. CONT., PROJ. NOS.
CARD 4. TITLE AND VOLUME.
CARD 5. PERSONAL AUTHORS.
CARD 6. SPECIAL DATA.
DATE (28-33)
PAGES (39-42)
COPY NOS. (49-65)
CLASS (69-71)
CARD 7. SUBJECTS.
CARD 8. TRACINGS.
AMCDO FORM 476 (one copy)
(REV) 20 AUG 1964
LIBRARY CATALOGING WORKSHEET

Figure E